Pattern Recognition in Spaces of Probability Distributions for the Analysis of Edge-Localized Modes in Tokamak Plasmas

Patroonherkenning in ruimtes van probabiliteitsdistributies voor de analyse van rand-gelocaliseerde modes in tokamakplasma's

Aqsa Shabbir

Promotors: Prof. G. Verdoolaege, PhD, Prof. H. Zohm, PhD
Doctoral thesis submitted in order to obtain the academic degrees of
Doctor of Engineering Physics (Ghent University) and
Doctor rerum naturalium (Ludwig-Maximilians-Universität München)

Department of Applied Physics
Head of Department: Prof. C. Leys, PhD
Faculty of Engineering and Architecture

UNIVERSITEIT GENT

Department of Physics
Head of Department: Prof. R. Bender, PhD
Faculty of Physics

Academic year 2015 - 2016

ISBN 978-90-8578-923-9
NUR 984, 926
Wettelijk depot: D/2016/10.500/55

This research work has been performed at:

Ghent University
Faculty of Engineering and Architecture
Department of Applied Physics
Sint-Pietersnieuwstraat 41
B-9000 Ghent, Belgium

Max-Planck-Institut für Plasmaphysik
Boltzmannstraße 2
D-85748 Garching, Germany

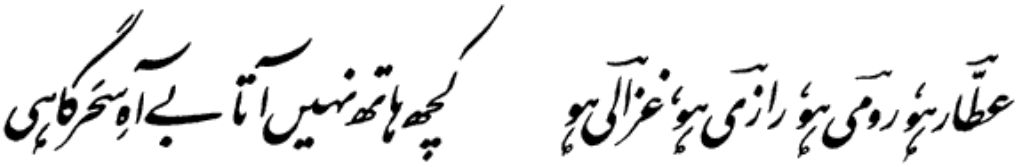
Joint European Torus
Culham Centre for Fusion Energy (CCFE)
Abingdon OX14 3DB, United Kingdom

Members of the examination committee

Chairman:	**Prof. Dr. Rik Van de Walle** *Ghent University*
Secretary:	**Prof. Dr. Jean-Marie Noterdaeme** *Ghent University* and *Max-Planck-Institut für Plasmaphysik*
Promoters:	**Prof. Dr. Geert Verdoolaege** *Ghent University* and *Royal Military Academy, Brussels* **Prof. Dr. Hartmut Zohm** *Max-Planck-Institut für Plasmaphysik*
Other members:	**Prof. Dr. Ralf Bender** *Ludwig-Maximilians University of Munich* **Prof. Dr. Otmar Biebel** *Ludwig-Maximilians University of Munich* **Prof. Dr. Gert de Cooman** *Ghent University* **Prof. Dr. Harald Lesch** *Ludwig-Maximilians University of Munich* **Prof. Dr. Gregor Morfill** *Ludwig-Maximilians University of Munich* and *terraplasma GmbH* **Dr. Andrea Murari** *Culham Science Centre* and *Consorzio RFX (CNR, ENEA, INFN, Acciaierie Venete S.P.A., Università di Padova)*

Doctoral guidance committee

Promoters:	**Prof. Dr. Geert Verdoolaege** *Ghent University* and *Royal Military Academy, Brussels* **Prof. Dr. Hartmut Zohm** *Max-Planck-Institut für Plasmaphysik*
Supervisor:	**Dr. Otto J.W.F Kardaun** *Max-Planck-Institut für Plasmaphysik*
Other members:	**Prof. Dr. Jean-Marie Noterdaeme** *Ghent University* and *Max-Planck-Institut für Plasmaphysik* **Dr. Andrea Murari** *Culham Science Centre* and *Consorzio RFX (CNR, ENEA, INFN, Acciaierie Venete S.P.A., Università di Padova)*



Like Attar, Rumi, Razi and Ghazali, One may be mystic, great or wise,
But none can reach his goal without hard work and morning sighs

Dr. Allama Muhammad Iqbal

Dedicated to my parents

Dr. Shahnaz Cheema & Dr. Shabbir Ahmed

for being my inspiration, for teaching me the importance of being human, for their sacrificial unconditional love & for encouraging me at each step of life,

especially this one

Acknowledgements

This doctoral thesis would be devoid of its spirit, if it had not been for the invaluable academic, educational, psychological and emotional contribution of several people, whom I would like to thank most sincerely.

However, first and foremost I humbly thank Allah Almighty for his countless blessings and for bestowing me with the strength and perseverance required for this endeavor.

Next I wish to express heartfelt gratitude for my joint PhD promoters, Prof. Geert Verdoolaege and Prof. Hartmut Zohm. Geert has led by example. He was always available for my questions and gave generously of his time and vast knowledge. It is hard to overstate my appreciation for him, except that this journey would have been impossible without him. Prof. Zohm has been an inspiration and I thank him profoundly for his guidance, enthusiasm and encouragement.

My sincere thanks also goes to Dr.Otto Kardaun for his immense knowledge, didactic guidance, patience and motivation. He inspired me with his tireless passion for the field and the stimulating and engaging discussions gave me the impetus to go further during the grueling work days.

I would like to express my appreciation and gratitude for Prof. Jean-Marie Noterdaeme and Prof. Guido Van Oost for their support and advice. My special thanks go to Dr. Gregoire Hornung and Frank Jansens. Gregoire is an amazing colleague and I have benefited greatly from many of his formal as well as informal advice. Frank, with his vivaciousness, enabled me to truly appreciate Belgian warmth and hospitality. Special thanks also goes to Kathleen Van Oost for ever so meticulously manging the bureaucratic aspects of my PhD. I would also like to acknowledge my officemates, Patrick Vanraes, and Xiaolong Deng for the many laughters and the many ups and downs which we faced together. A special mention and gratitude is again due for Geert, Gregoire, Frank and Kathleen for the many

lunches we shared, the fond memories of which will always remain with me.

It is imperative to thank all colleagues at the Department of Applied Physics, Ghent University, Max-Planck Institute for Plasma Physics (IPP) and Culham Center for Fusion Energy (CCFE-JET) as well as all the fellow FUSION-DC PhD students, for contributing and becoming a unique part of my PhD journey.

I seize this opportunity to also acknowledge and thank my colleagues and superiors at Lahore College for Women University, Lahore and my teachers and classmates at University of Engineering and Technology, Lahore for contributing to my growth, inspiring me and providing me with invaluable experiences.

Finally, I cannot agree more with a wise man who once said: family is not an important thing, it is everything. I want to humbly thank my late maternal grandparents, Chahudhry Rehmatullah Cheema and Afzal Begum, who were true visionaries and who laid the foundation for a series of strong, empowered and educated women. No words can describe my gratitude for my parents Dr. Shahnaz Cheema and Dr. Shabbir Ahmed. *Mama , Papa*, I owe it all to you. Next, I want to thank my best friend (my *jiggy*) Yasmin Ansari for understanding me, accepting me, supporting me and for being my sister, that I never had. I am truly fortunate to have you in my life.

Moving on to the most special corner of my heart, I want to thank my daughter, my princess and my little fairy, Fatima Ali, for being my strength and my biggest motivation. Next, I want to thank my son, Mustafa Ali, who accompanied me in my womb as I attended conferences, worked on manuscripts and conducted experiments. Thank you for being our little bundle of joy. And, finally I thank my husband, Ali Hussain Kazim, for his patience, love, friendship, humor and support. Having you as my husband, is a constant reminder that Allah is happy with me. Thanks for everything!

Aqsa Shabbir

En route Ghent to Munich

22 April 2016

Abstracts

English summary

Magnetically confined fusion plasmas provide several data analysis challenges due to the occurrence of massive data sets, substantial measurement uncertainty, stochasticity and data dimensionality, and often nonlinear interactions between measured quantities. Recently, methods from the fields of machine learning and probability theory—some standard, some more advanced—have come to play an increasingly important role in analyzing data from fusion experiments. The capabilities offered by such methods to efficiently extract, possibly in real time, additional information from the data that is not immediately apparent to human experts, has attracted attention from an increasing number of researchers. In addition, innovative methods for real-time data processing can play an important role in plasma control, in order to ensure safe and reliable operation of the machine.

Pattern recognition is a discipline within the information sciences that concerns the exploration of structure in (multidimensional) data sets using computer-based methods and algorithms. In this doctoral work, pattern recognition techniques are developed and applied to data from tokamak plasmas, in order to contribute to a systematic analysis of *edge-localized modes* (ELMs). ELMs are magnetohydrodynamic (MHD) instabilities occurring in the edge region of high-confinement (H-mode) fusion plasmas. The type I ELMy H-mode is the reference scenario for operation of the next-step fusion device ITER. On the one hand, ELMs have a beneficial effect on plasma operation through their role in impurity control. On the other hand, ELMs eject energy and particles from the plasma and, in ITER, large unmitigated ELMs are expected to cause intolerable heat loads on the plasma-facing components (PFCs).

In interpreting experiments focused on ELM understanding and control, a

significant challenge lies in handling the measurement uncertainties and the inherent stochasticity of ELM properties. In this work, we employ probabilistic models (distributions) for a quantitative data description geared towards an enhanced systematization of ELM phenomenology. Hence, we start from the point of view that the fundamental object resulting from the observation of a system is a probability distribution, with every single measurement providing a sample from this distribution. We argue that, particularly for richly stochastic phenomena like ELMs, the probability distribution of physical quantities contain significantly more information compared to mere averages. Consequently, in exploring the patterns emerging from the various ELM regimes and relations, we need methods that can handle the intrinsic probabilistic nature of the data.

The original contributions of this work are twofold. First, several novel pattern recognition methods in non-Euclidean spaces of probability distribution functions (PDFs) are developed and validated. The second main contribution lies in the application of these and other techniques to a systematic analysis of ELMs in tokamak plasmas.

In regard to the methodological aims of the work, we employ the framework of *information geometry* to develop pattern visualization and classification methods in spaces of probability distributions. In information geometry, a family of probability distributions is considered as a Riemannian manifold. Every point on the manifold represents a single PDF and the distribution parameters provide local coordinates on the manifold. The Fisher information plays the role of a Riemannian metric tensor, enabling calculation of geodesic curves on the surface. The length of such curves yields the *geodesic distance* (GD) on probabilistic manifolds, which is a natural similarity (distance) measure between PDFs. Equipped with a suitable distance measure, we extrapolate several distance-based pattern recognition methods to the manifold setting. This includes k-nearest neighbor (kNN) and conformal predictor (CP) methods for classification, as well as multidimensional scaling (MDS) and landmark multidimensional scaling (LMDS) for data visualization (dimensionality reduction). Furthermore, two new classification schemes are developed: a distance-to-centroid classifier (D2C) and a principal geodesic classifier (PGC). D2C classifies on the basis of the minimum GD to the class centroids and PGC considers the shape of the class on the manifold by determining the minimum

distance to the principal geodesic of each class. The methods are validated by their application to the classification and retrieval of colored texture images represented in the wavelet domain. Both methods prove to be computationally efficient, yield high accuracy and also clearly exhibit the adequacy of the GD and its superiority over the Euclidean distance, for comparing PDFs. This also aids in demonstrating the utility and adaptability of the developed methods to a wide range of applications other than ELMs, which are the prime focus of analysis in this work.

The second main goal of the work targets ELM analysis at three fronts, using pattern recognition and probabilistic modeling :

(i) We first concentrate on *visualization* of ELM characteristics by creating maps containing projections of multidimensional ELM data, as well as the corresponding probabilistic models. Such maps can provide physicists and machine operators with a convenient means and a useful tool for plasma monitoring and for studying data patterns reflecting key regimes and their underlying physics. In particular, GD-based MDS is used for representing the complete distributions of the multidimensional data characterizing the operational space of ELMs onto two-dimensional maps. Clusters corresponding to type I and type III ELMs are identified and the maps enable tracking of trends in plasma parameters across the operational space. It is shown that the maps can also be used with reasonable accuracy for predicting the values of the plasma parameters at a certain point in the operational space.

(ii) Our second application concerns fast, standardized and automated *classification* of ELM types. ELM types have so far been identified and characterized on an empirical and phenomenological basis. The presented classification schemes are aimed at complementing the phenomenological characterization using standardized methods that are less susceptible to subjective interpretation, while considerably reducing the effort of ELM experts in identifying ELM types. To this end, different classification paradigms (parametric and non-parametric) are explored and put to use. Discriminant analysis (DA) is used for determining a linear separation boundary between type I and III ELMs in terms of global plasma parameters, which can then be used for the prediction of ELM types as well as the

study of ELM occurrence boundaries and ELM physics. However, DA makes an assumption about the underlying class distribution and presently cannot be applied in spaces of probability distributions, leading to a suboptimal treatment of stochasticity. This is circumvented by the use of GD-based CP and kNN classifiers. CP provides estimates of its own accuracy and reliability and kNN is a simple, yet powerful classifier of ELM types. It is shown that a classification based on the distribution of ELM properties, namely inter-ELM time intervals and the distribution of global plasma parameters, is more informative and accurate than the classification based on average parameter values.

(iii) Finally, the *correlation* between ELM energy loss (ELM size) and ELM waiting times (inverse ELM frequency) is studied for individual ELMs in a set of plasmas from the JET tokamak upgraded with the ITER-like wall (ILW). Typically, ELM control methods rely on the empirically observed inverse dependence of average ELM energy loss on average ELM frequency, even though ELM control is targeted at reducing the size of individual ELMs and not the average ELM loss. The analysis finds that for individual ELMs the correlation between ELM energy loss (W_{ELM}) and waiting times (Δt_{ELM}) varies from zero to a moderately positive value. A comparison is made with the results from a set of carbon-wall (CW) JET plasmas and nitrogen-seeded ILW JET plasmas. It is found that a high correlation between W_{ELM} and Δt_{ELM} comparable to CW plasmas is only found in nitrogen-seeded ILW plasmas. Furthermore, most of the unseeded JET ILW plasmas have ELMs that are followed by a second phase referred to as the *slow transport event* (STE). The effect of the STEs on the distribution of ELM durations is studied, as well as their influence on the correlation between W_{ELM} and Δt_{ELM}. This analysis has a clear outcome for the optimization of ELM control methods, while presenting insights for an improved physics understanding of ELMs.

Nederlandse samenvatting

In de context van magnetisch opgesloten fusieplasma's bestaan er verschillende uitdagingen op het gebied van data-analyse, vanwege de grootte van de datasets, de aanzienlijke meetonzekerheden, stochasticiteit en dimensionaliteit, en de vaak niet-lineaire interacties tussen de gemeten grootheden. Methodes uit het domein van machinaal leren en probabiliteitstheorie—sommige standaard, andere meer geavanceerd—spelen sinds kort een steeds belangrijkere rol in de analyse van data van fusie-experimenten. De mogelijkheden die dergelijke methodes bieden om, eventueel in reële tijd, op een efficiënte manier bijkomende informatie uit de data te halen, die niet onmiddellijk in het oog springt van menselijke experts, heeft de aandacht getrokken van een toenemend aantal onderzoekers. Bovendien kunnen innovatieve methodes voor real-time dataverwerking een belangrijke rol spelen in plasmacontrole, om veilige en betrouwbare operatie van de machine te verzekeren.

Patroonherkenning is een discipline binnen de informatiewetenschappen waarin structuur in (meerdimensionale) datasets bestudeerd wordt, gebruik makend van computergebaseerde methodes en algoritmes. In dit doctoraatswerk worden patroonherkenningstechnieken ontwikkeld en toegepast op data van tokamakplasma's, om zo bij te dragen tot een systematische analyse van *rand-gelocaliseerde modes* (edge-localized modes, of ELMs). ELMs zijn magnetohydrodynamische (MHD) instabiliteiten die voorkomen in de rand van fusieplasma's in het hoge-opsluitingsregime (H-mode). De type I-ELMige H-mode is het referentiescenario voor de operatie van ITER, de fusiemachine van de volgende generatie. Enerzijds hebben ELMs een gunstig effect op plasmaoperatie door hun rol in de controle van onzuiverheden. Anderzijds stoten ELMs energie en deeltjes uit het plasma en de verwachting is dat, in ITER, grote niet-getemperde ELMs een ontoelaatbaar hoge hittebelasting zullen veroorzaken op de wandcomponenten.

Bij de interpretatie van experimenten die focussen op de studie van ELMs en ELM-controle, ligt een grote uitdaging in een geschikte behandeling van meetonzekerheden en de inherente stochasticiteit van de eigenschappen van ELMs. In dit werk gebruiken we probabilistische modellen (distributies) voor een kwantitatieve beschrijving van de data met het oog op een verbeterde systematisering van ELM-fenomenologie. We gaan dus uit van het standpunt dat een

probabiliteitsdistributie het fundamentele object is dat voortkomt uit de observatie van een systeem, waarbij elke individuele meting een sample voorstelt van deze verdeling. We betogen dat de probabiliteitsdistributie van fysische grootheden significant meer informatie bevat dan louter gemiddeldes, in het bijzonder voor fenomenen als ELMs met rijke stochastische karakteristieken. Bij de studie van patronen voortkomend uit de verschillende ELM-regimes en -relaties, hebben we bijgevolg methodes nodig die kunnen omgaan met de intrinsieke probabilistische natuur van de data.

De originele bijdragen van dit werk zijn tweeledig. Om te beginnen worden verscheidene nieuwe patroonherkenningsmethodes ontwikkeld en gevalideerd in niet-Euclidische ruimtes van probabiliteitsdistributies (probability distribution functions, of PDFs). Vervolgens worden deze en andere technieken toegepast voor een systematische analyse van ELMs in tokamakplasma's.

Met betrekking tot de methodologische doelstellingen van het werk, gebruiken we het raamwerk van de *informatiemeetkunde* om methodes te ontwikkelen voor visualisatie en classificatie in ruimtes van probabiliteitsdistributies. In de informatiemeetkunde wordt een familie van probabiliteitsdistributies beschouwd als een Riemanniaanse variëteit. Elk punt op de variëteit stelt één enkele PDF voor, en de parameters van de distributie dienen als lokale coördinaten op de variëteit. De Fisher-informatie speelt de rol van een Riemanniaanse metrische tensor, waarmee geodetische curves berekend kunnen worden op het oppervlak. De lengte van zulke curves geeft de *geodetische afstand* (geodesic distance, of GD) op probabilistische variëteiten, die een natuurlijke maat is van gelijkenis (afstand) tussen PDFs. Uitgerust met een geschikte afstandsmaat, zetten we verschillende afstandsgebaseerde patroonherkenningsmethodes om voor toepassing op variëteiten. Hieronder vallen de methode van de k-dichtste-buren (k-nearest neighbors, of kNN) en conforme predictors (CP) voor classificatie, alsook multidimensionale schaling (multidimensional scaling, of MDS) en MDS met oriëntatiepunten voor het visualiseren van data (dimensionaliteitsreductie). Verder worden twee nieuwe classificatiemethodes ontwikkeld: afstand-tot-centroïde (distance-to-centroid, of D2C) en principale geodetische classificatie (principal geodesic classification, of PGC). D2C classificeert op basis van de minimale geodetische afstand tot de klassecentroïdes en PGC brengt de vorm van de klasse op de variëteit in rekening

door de minimale afstand te bepalen tot de principale geodeet van elke klasse. De methodes worden gevalideerd door toepassing op de classificatie en het opzoeken van beelden met gekleurde texturen, voorgesteld in het waveletdomein. Beide methodes blijken computationeel efficiënt, hebben een hoge nauwkeurigheid en geven ook duidelijk aan dat de GD geschikt is en superieur is aan de Euclidische afstand om PDFs te vergelijken. Op deze manier wordt ook het nut aangetoond van de ontwikkelde methodes in een hele reeks toepassingen buiten ELMs, welke het zwaartepunt vormen van de analyse in dit werk, en de mogelijkheid om de methodes hiernaar aan te passen.

De tweede hoofddoelstelling van dit werk beoogt de analyse van ELMs op drie fronten, gebruik makend van patroonherkenning en probabilistische modellering:

(i) We richten ons eerst op de *visualisatie* van ELM-karakteristieken door kaarten te creëren die de projectie tonen van meerdimensionale ELM-data, en van de corresponderende probabilistische modellen. Dergelijke kaarten voorzien fysici en machineoperatoren van een handige manier en een nuttig instrument voor het volgen van plasma's en het bestuderen van patronen in de data die belangrijke regimes weergeven en hun onderliggende fysica. Meer in het bijzonder wordt GD-gebaseerde MDS gebruikt om in tweedimensionale kaarten de complete distributies voor te stellen van meerdimensionale data die de operationele ruimte karakteriseert van ELMs. Clusters corresponderend met type I- en type III-ELMs worden geïdentificeerd en de kaarten laten toe om trends te volgen in de plasmaparameters doorheen de operationele ruimte. We tonen aan dat de kaarten ook gebruikt kunnen worden, met redelijke nauwkeurigheid, om de waarden van de plasmaparameters te voorspellen in een bepaald punt in de operationele ruimte.

(ii) Onze tweede toepassing betreft snelle, gestandaardiseerde en geautomatiseerde *classificatie* van ELM-types. Tot nu toe werden ELM-types geïdentificeerd en gekarakteriseerd op een empirische en fenomenologische basis. De voorgestelde classificatiemethodes zijn erop gericht om de fenomenologische karakterisering te vervolledigen door middel van gestandaardiseerde methodes die minder onderworpen zijn aan subjectieve interpretatie, terwijl de werklast voor ELM-experten gevoelig gereduceerd wordt bij het identificeren

van ELM-types. Hiertoe worden verschillende classificatieparadigma's (parametrisch en niet-parametrisch) getest en ingezet. Discriminantanalyse (DA) wordt gebruikt om een lineaire grens tussen de type I- en type III-ELMs vast te leggen in termen van globale plasmaparameters, die vervolgens gebruikt kan worden om ELM-types te voorspellen en voor de studie van de operationele regio's waar ELMs voorkomen en de fysica van ELMs. DA maakt evenwel een veronderstelling met betrekking tot de onderliggende klassedistributie en momenteel kan deze methode niet toegepast worden in ruimtes van probabiliteitsdistributies, met als gevolg een suboptimale behandeling van stochasticiteit. Dit wordt verholpen door GD-gebaseerde CP- en kNN-classificatie te gebruiken. CP geeft een schatting van zijn eigen nauwkeurigheid en betrouwbaarheid en kNN is een eenvoudige maar krachtige classificatiemethode voor ELM-types. We tonen aan dat classificatie op basis van de distributie van ELM-eigenschappen, namelijk de inter-ELM-tijd en globale plasmaparameters, meer informatief en nauwkeuriger is dan classificatie met behulp van enkel gemiddelde waardes van parameters.

(iii) Ten slotte wordt de *correlatie* tussen het energieverlies van ELMs (ELM-grootte) en de ELM-wachttijd (inverse van de ELM-frequentie) bestudeerd voor individuele ELMs in een set plasma's van de JET-tokamak uitgerust met de ITER-achtige wand (ITER-like wall, of ILW). Controlemethodes voor ELMs berusten typisch op de empirisch geobserveerde inverse afhankelijkheid tussen het gemiddelde energieverlies en de gemiddelde frequentie van ELMs, hoewel ELM-controle eigenlijk beoogt de grootte van individuele ELMs te reduceren en niet het gemiddelde ELM-verlies. Uit de analyse blijkt dat voor individuele ELMs de correlatie tussen het energieverlies (W_{ELM}) en de wachttijd (Δt_{ELM}) varieert tussen nul en matig. Een vergelijking wordt gemaakt tussen de resultaten in een reeks JET-plasmas met de koolstofvezelwand (carbon wall, of CW) en JET-ILW-plasma's met toegevoegde stikstof. Hieruit blijkt dat een hoge correlatie tussen W_{ELM} en Δt_{ELM}, vergelijkbaar met CW-plasma's, enkel teruggevonden wordt in ILW-plasma's met toegevoegde stikstof. Bovendien worden de ELMs in de meeste JET-ILW-plasma's zonder stikstof gevolgd door een tweede fase genaamd het *trage-transport-verschijnsel* (slow transport event, of STE). Het

effect van de STEs op de distributie van de ELM-duur wordt onderzocht, samen met hun invloed op de correlatie tussen W_{ELM} en Δt_{ELM}. De analyse heeft een duidelijk gevolg voor de optimalisatie van controlemethodes voor ELMs en zorgen voor bijkomend inzicht in de fysica van ELMs.

Deutsche Zusammenfassung

Die Analyse von experimentellen Daten magnetisch eingeschlossener Fusionsplasmen stellt wegen der großen Datenmengen, der hohen Dimensionalität, der Messunsicherheiten und auch der oft nichtlinearen Beziehungen untereinander eine große Herausforderung dar. Methoden der Datenanalyse aus den Feldern des maschinellen Lernens sowie der Wahrscheinlichkeitstheorie spielen daher in letzter Zeit eine immer größere Rolle bei der Analyse von Daten aus Fusionsexperimenten. Dabei interessiert vor allem die Möglichkeit, zusätzliche Information welche dem menschlichen Beobachter verborgen bleiben, systematisch zu extrahieren. Zusätzlich können innovative Methoden der Echtzeit-Datenverarbeitung eine wichtige Rolle für Kontrollanwendungen in Fusionsexperimenten spielen.

Mustererkennung ist eine Disziplin der Informationstheorie welche sich mit der Erforschung von Strukturen in multidimensionalen Datensätzen durch computergestützte Methoden und Algorithmen beschäftigt. In dieser Doktorarbeit werden Methoden der Mustererkennung auf Daten von Tokamakexperimenten für eine systematische Analyse von edge-localized modes (ELMs) angewendet. ELMs sind magnetohydrodynamische (MHD) Instabilitäten die am Plasmarand in 'high-confinement' (H-mode) Fusionsplasmen auftreten. Die 'Typ I ELMy H-mode' ist das Referenz-Betriebsszenario für das zukünftige ITER Experiment. ELMs spielen einerseits eine positive Rolle für den Plasmabetrieb da sie zur Verunreinigungskontrolle beitragen. Andererseits werfen ELMs Teilchen und Energie aus dem Plasma und könnten daher in ITER die Integrität der ersten Wand gefährden.

Eine signifikante Herausforderung bei der Interpretation von Experimenten welche sich mit dem Verständnis und der Kontrolle von ELMs beschäftigen liegt in der Behandlung der Messunsicherheiten sowie der inhärenten Stochastizität der ELM Parameter. In der vorliegenden Arbeit werden probabilistische Modelle (Verteilungen) zur quantitativen Beschreibung der Daten mit dem Ziel einer verbesserten systematischen Einteilung der ELM-Phänomenologie verwendet. Dabei wird davon ausgegangen, dass die fundamentale Größe eines Systems eine Wahrscheinlichkeitsverteilung ist, wobei jede Einzelmessung eine Stichprobe dieser Verteilung darstellt. Dabei wird angenommen dass, im Besonderen für

stark stochastische Ereignisse wie ELMs, die Wahrscheinlichkeitsverteilung der physikalischen Parameter deutlich mehr Information enthält als deren Mittelwerte. Folglich erfordert die Erforschung der Struktur der unterschiedlichen ELM Regimes Methoden, welche die intrinsisch stochastische Natur der Daten berücksichtigen kann.

Diese Arbeit liefert zwei grundsätzlich neue Beiträge: zunächst werden neuartige Strukturerkennungs-Methoden in nicht-euklidischen Räumen von Wahrscheinlichkeitsverteilungen entwickelt und validiert. Der zweite grundsätzliche Beitrag liegt in der Anwendung dieser und anderer Methoden auf eine systematische Analyse von ELMs in Tokamakplasmen.

Aus methodologischer Sicht wird in dieser Arbeit die Informationsgeometrie angewendet um Methoden zur Mustererkennung und –klassifizierung in Räumen von Wahrscheinlichkeitsverteilungen zu entwickeln. In der Informationsgeometrie wird eine Familie von Wahrscheinlichkeitsverteilungen als eine Riemannsche Mannigfaltigkeit aufgefasst. Jeder Punkt auf der Mannigfaltigkeit stellt eine Wahrscheinlichkeitsverteilung dar und die Verteilungsparameter sind lokale Koordinaten auf der Mannigfaltigkeit. Die Fisher Information spielt dabei die Rolle des Riemannschen metrischen Tensors und erlaubt es, geodätische Kurven auf der Fläche zu berechnen. Die Länge einer solchen Kurve ergibt den geodätischen Abstand auf der Mannigfaltigkeit, welcher ein natürliches Maßfür den Abstand zwischen Verteilungsfunktionen ist. Mit diesem geeigneten Abstandsmaßwerden mehrere Mustererkennungsmethoden welche auf dem Abstand basieren auf die Mannigfaltigkeit angewandt. Diese schließen die 'k-nearest neighbor' (kNN) und 'conformal predictor' (CP) Klassifikationsmethoden ein sowie 'multidimensional scaling' (MDS) und 'landmark multidimensional scaling' (LMDS) zur Datenvisualisierung mit dem Ziel der Dimensionsreduktion. Desweitern werden zwei neue Klassifikationsmethoden entwickelt: ein 'distance-to-centroid classifier' (D2C) und ein 'principal geodesic classifier' (PGC). D2C klassifiziert auf Basis des minimalen geodätischen Abstands vom Schwerpunkt der Daten und PGC berücksichtigt die Form der Klasse auf der Mannigfaltigkeit indem der Abstand zur Hauptgeodätischen jeder Klasse bestimmt wird. Diese Methoden werden durch Anwendung auf die Klassifizierung und Rekonstruktion von farbigen Texturbildern in der Waveletdarstellung validiert. Beide Methoden stellen sich als effizient im

Rechenaufwand heraus und liefern hohe Genauigkeit, wobei der geodätische Abstand dem euklidischen Abstand deutlich überlegen ist und somit als angemessen für den Vergleich von Verteilungsfunktionen bestätigt wird. Dies dient auch dem Nachweis der Eignung der entwickelten Methoden für eine Vielzahl von Anwendungen über das in dieser Arbeit vorrangig behandelte Feld der ELMs hinaus.

Das zweite Hauptziel der Arbeit ist die Analyse von ELMs mit den Methoden der Mustererkennung und der wahrscheinlichkeitstheoretischen Modellierung auf drei Gebieten:

(i) Zunächst wird die Visualisierung von ELM Eigenschaften durch Erstellung von Abbildungen behandelt welche multidimensionale ELM Daten projizieren. Solche Abbildungen können für Physiker und Experimentatoren ein nützliches Werkzeug zur Überwachung der Plasmaentladung darstellen und dienen darüber hinaus zu Studien von Datenmustern, welche prinzipielle Regimes und deren zugrundeliegende Physik charakterisieren. Im speziellen wird die GD-basierte MDS zur Darstellung der gesamten Verteilung der multidimensionalen Daten, welche das Auftreten von ELMs beschreiben in zweidimensionalen Abbildungen verwendet. Cluster in welchen 'Typ I' und 'Type III' ELMs auftreten werden identifiziert und die Abbildung ermöglicht es, Trends in der Veränderung von Plasmaparametern im Parameterraum zu erkennen. Es wird gezeigt, dass diese Abbildungen auch dazu verwendet werden können, die Plasmaparameter für einen bestimmten Punkt im Betriebsbereich vorherzusagen.

(ii) Eine zweite Anwendung beschäftigt sich mit einer schnellen, standardisierten Klassifizierung des ELM Typs. ELM Typen wurden bisher auf einer empirisch-phänomenologischen Basis identifiziert. Die hier vorgestellten Klassifizierungs-Schemata dienen der Ergänzung der phänomenologischen Beschreibung durch standardisierte Methoden welche weniger anfällig für subjektive Wahrnehmung und Interpretation sind und sollen auch den Aufwand bei der Bestimmung des ELM Typs verringern. Verschiedene Klassifizierungsmethoden, parametrisch und nicht-parametrisch, werden untersucht und eingesetzt. Discriminant Analysis (DA) wird für die Bestimmung einer linearen Grenze zwischen Typ I und Typ III ELMs

in globalen Plasmaparametern eingesetzt, die dann sowohl zur Vorhersage des ELM Typs als auch zur Untersuchung der Bereiche, in denen die unterschiedlichen ELM Typen auftreten, verwendet wird. Dabei basiert die DA allerdings auf einer Annahme über die zugrunde liegende Verteilung der Klassen und kann nach derzeitigem Stand nicht auf Räume von Verteilungsfunktionen angewendet werden, was zu einer unzureichenden Behandlung der Stochastizität führt. Dies wird durch die Verwendung von GD-basierter CP und von kNN Klassifikatoren behoben. CP liefert eine Abschätzung ihrer Genauigkeit und Zuverlässigkeit und kNN ist ein einfacher, aber leistungsstarker Klassifikator für ELM-Typen. Es wird gezeigt dass eine Klassifizierung basierend auf der Verteilung der ELM Eigenschaften, namentlich der inter-ELM Zeitintervalle und der Verteilung der globalen Plasmaparameter, mehr Information enthält als eine Klassifizierung welche auf gemittelten Werten basiert.

(iii) Schließlich wird die Korrelation zwischen ELM Energieverlust (ELM Größe) und ELM Wartezeiten (inverse ELM Frequenz) für individuelle ELMs aus einer Datenbasis von Plasmaentladungen des JET Tokamaks in der 'ITER-like wall' (ILW) Konfiguration untersucht. ELM Kontrollmethoden basieren typischerweise auf dem empirisch beobachteten inversen Zusammenhang zwischen mittlerem ELM-Verlust und mittlerer ELM-Frequenz, obwohl ELM Kontrolle die Reduktion der Größe individueller ELMs zum Ziel hat. Die Analyse zeigt, dass für individuelle ELMs die Korrelation zwischen ELM-Energieverlust (W_{ELM}) und Wartezeit (Δt_{ELM}) generell niedrig ist. Dieses Ergebnis wird mit einem Datensatz von JET in der 'carbon-wall' (CW) Konfiguration sowie einem Datensatz von Stickstoff-gekühlten ILW JET Plasmen verglichen. Es zeigt sich, dass eine hohe Korrelation zwischen W_{ELM} und Δt_{ELM}, vergleichbar zu CW Plasmen, nur in Stickstoff-gekühlten ILW Plasmen auftritt. Darüber hinaus treten in den meisten JET ILW Plasmen ohne Stickstoffkühlung ELMs auf, welche von einer zweiten Phase, slow transport event (STE) genannt, begleitet werden. Der Effekt der STEs auf die Verteilung der ELM Dauer sowie deren Einfluss auf die Korrelation zwischen W_{ELM} und Δt_{ELM} wird untersucht. Diese Untersuchung hat einerseits eine starke Relevanz für die Optimierung von Methoden zur ELM Kontrolle,

andererseits trägt sie zum tieferen Einblick in die den ELMs zugrunde liegende Physik bei.

Contents

List of Figures

List of Tables

Acronyms

A

Alcator C-Mod Alto Campo Toro, High Field Torus. A tokamak operational in MIT, Cambridge, USA.

AM Ambiguity (%)

ASDEX AXially Symmetric Divertor EXperiment. Medium sized tokamak operational from 1983-1990 in Garching, Germany.

AUG ASDEX Upgrade (AXially Symmetric Divertor EXperiment). Medium sized tokamak operational in Garching, Germany.

AUG ELM-DBI A database of 30 full-W wall AUG plasmas, reported in A.2

C

CFC Carbon fibre composite

CO Average confidence (%)

CP Conformal predictor classifier

CR Average credibility (%)

CW Carbon wall at JET tokamak

D

D2C Distance-to-centroid classifier

DIII-D Doublet III-D. Medium sized tokamak operational in San Diego, USA.

DEMO DEMOnstration power plant. Proposed nuclear fusion power station.

E

EAST Experimental Advanced Superconducting Tokamak. A superconducting tokamak operational in Heifei, China.

ELM Edge-localized mode

ER Error rate (%)

ETB Edge transport barrier

G

GD Rao geodesic distance

GGD Generalized Gaussian distribution

H

H-mode High confinement mode

HFS High field side

I

ICP Inductive conformal predictor

ICRH Ion cyclotron resonance heating

ILW ITER-like wall at JET tokamak

ITB Internal transport barrier

ITER International Thermonuclear Experimental Reactor. World's largest tokamak, currently under construction in Southern France.

ITPA International Tokamak Physics Activity

ITPA database ITPA global H-mode confinement database (DB3, version 13f)

J

JET Joint European Torus. World's largest operational tokamak, situated in Culham, Oxfordshire, UK.

JET ELM-DBI A database of 100 CW JET plasmas, reported in A.1

JET ELM-DBII A database of 38 ILW and 20 CW JET plasmas, reported in A.3

JFT-2M JAERI Fusion Torus-2M. A medium size tokamak operational from 1983-2004 in Naka, Japan.

JT-60U Japan Torus-60U. A tokamak operational from 1985-2010 in Naka, Japan.

K

KLD Kullback-Leibler divergence

kNN k-nearest neighbor classifier

L

L-mode Low confinement mode

LDA Linear discriminant analysis

LFS Low field side

LMDS Landmark multidimensional scaling

M

MAST Mega Amperical Spherical Tokamak. A spherical tokamak, operational from 1999-2013 in Culham, Oxfordshire, UK.

MDS Multidimensional scaling

MGGD Multivariate generalized Gaussian distribution

MHD Magnetohydrodynamics. Fluid description of magnetized plasma.

MLM Machine learning method

N

NBI Neutral beam injection heating

NSTX National Spherical Torus Experiment. A spherical tokamak, operational from 1999-2012 in Princeton, New Jersey, US.

P

PCA Principal component analysis

PDF Probability distribution function, may also refer to a probability density function

PFC Plasma facing component

PGA Principal geodesic analysis

PGC Principal geodesic classifier

Q

QDA Quadratic discriminant analysis

S

SOL Scrape-off Layer. Region between confined plasma and wall.

SR Success rate (%)

STE Slow transport event

STD Standard deviation

T

TCP Transductive conformal predictor

Chapter 1

Introduction

1.1 Energy production context

The continual rise in global energy demand, geopolitical instability and heightened concerns for climate change are unanimously pressing for clean, safe and sustainable sources of energy. World energy consumption, fueled by expanding global economies and population growth, is expected to quadruple by 2100 [1]. On the other hand, world's energy production is faced with multi-pronged challenges. As shown in Figure 1.1 fossil fuels (oil, natural gas and coal) in the year 2014 accounted for a whopping 87% of world's total energy consumption. Not only do they pose climate change concerns but they are also subject to volatile geopolitical stability. At the recent UN climate change conference (COP21/CMP11) held in Paris [2], representatives from 195 countries, acknowledged the impending dangers of climate change and resolved to limit global warming to "well-below" 2°C above pre-industrial levels. This in turn implies that the world needs to move rapidly towards zero net carbon emissions.

Renewable energy sources including hydroelectricity account for less than 10% of the total energy consumption. Despite a consensus that renewable energy sources are the only long term solution, wind and solar power are limited by two major constraints: circumstantial availability and low power density. While energy produced in nuclear power plants using fission reactions does not suffer from these bottlenecks, the risks associated with a nuclear accident and the long term storage of radioactive waste, makes this energy source increasingly unpopular. Further, the 2011 Fukushima Daiichi nuclear disaster caused deep public anxiety throughout the

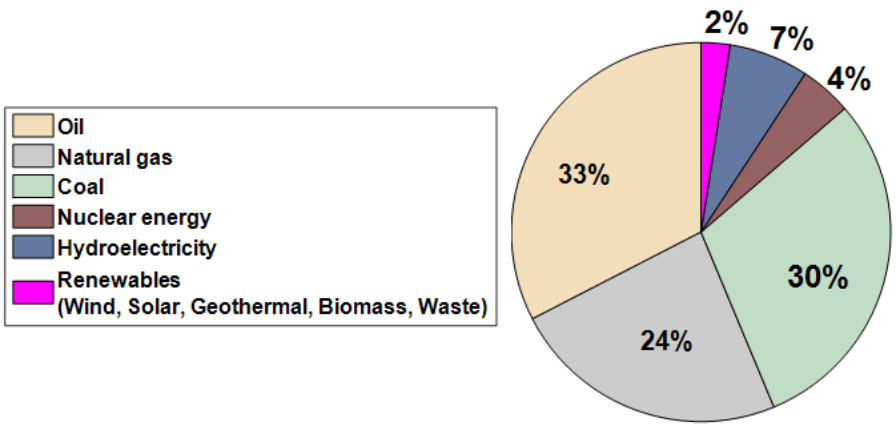


Figure 1.1: *World's primary energy consumption in 2014. Fossil fuels account for 87% of the total energy consumption.*

world and has led many countries to accelerate systematic phase out of nuclear power production. Given this, *nuclear fusion*, also based on nuclear reactions but devoid of certain risks associated with fission, is another promising approach.

Nuclear fusion is the energy source that powers the stars and the sun. The final goal of scientific research on controlled thermonuclear fusion is to produce energy by fusion in a power plant on earth. The following advantages make nuclear fusion a worthwhile pursuit:

- *Sustainable*: Fusion fuels: deuterium (D) and tritium (T), are widely available and nearly inexhaustible. Deuterium has a natural abundance of approximately one atom in 6400 hydrogen atoms and can be distilled from all forms of water. Tritium, on the other hand, will be bred during the fusion reaction in a blanket containing lithium. Available quantities of lithium and deuterium in nature are sufficient to cover global energy consumption for a million years [3].

- *Clean*: Unlike fossil fuels, no greenhouse emissions occur. The major by-product is helium which is a non-toxic inert gas.

- *Safe*: Unlike nuclear fission there will be no long-lived radioactive wastes. Radioactive waste will mainly come from the reactor walls and will have a lifetime of less than 100 years. Secondly, there is no risk of a Fukushima type nuclear accident. In the wake of a disturbance, fusion fuel cools down within seconds, averting the onset of a chain reaction and subsequent reactor melt down.

- *Abundant*: Energy released via nuclear fusion will be four million times more than a chemical reaction such as the burning of fossil fuels and four times as much as nuclear fission reactions at equal mass [4].

1.2 Nuclear fusion

Nuclear fusion occurs when two light nuclei collide at sufficiently high energy required to overcome the repulsive Coulomb forces and become subject to the strong interaction forces at short distances ($10^{-15}m$). The resulting nucleus has a lower total mass than the sum of the two original nuclei and the mass deficit Δm is converted into energy through $E = \Delta mc^2$ [5]. The most promising reaction for nuclear fusion on earth is between D and T, isotopes of hydrogen.

$$ {}^{2}_{1}D + {}^{3}_{1}T \rightarrow {}^{4}_{2}He + {}^{1}n + 17.6\ MeV. \tag{1.1} $$

The energy released in the reaction is distributed between the kinetic energy of the α-particle ${}^{4}_{2}He$ $(3.5MeV)$ and the neutron n $(14.1MeV)$. At energies currently within reach, the $D-T$ reaction has the highest reaction cross-section compared to other fusion reactions.

It is only at extremely high temperatures that the nuclei possess enough energy to breach the Coulomb repulsion barrier. The average translational kinetic energy of a gas molecule at temperature T is $3/2k_BT$ where k_B is Boltzmann's constant. It follows from this that the Coulomb barrier can be overcome at a temperature of 3×10^9 K. Fortunately, due to the presence of a significant fraction of particles in the Maxwellian tail of the velocity distribution and quantum tunneling effects, the resulting temperature at which $D-T$ fusion can be achieved is about $10\,keV$. At such high temperatures the reactants are fully ionized and form a *plasma* state. Plasma as one of the four states of matter constitute most of the visible matter in the universe including Aurora Borealis and fluorescent lights on earth.

In addition to the required temperature, there are two further requirements for making the fusion reaction in equation (1.1) work: a minimum energy confinement time (τ_E) and an adequately high particle number density (n). As seen in Figure 1.2, in order to sustain an efficient fusion reaction, the requirements on T, n and τ_E

must be met simultaneously. In order to achieve self sustained cycle of fusion reactions, known as *ignition*, the triple product of T, n and τ_E must satisfy the Lawson criterion:

$$nT\tau_E > 5 \times 10^{21} \quad m^{-3}keVs. \tag{1.2}$$

High temperature vital for fusion prevents the use of a solid confinement structure. This is circumvented by the use of two main approaches: inertial confinement and magnetic confinement of the fusion plasma. Inertial confinement uses high plasma density ($10^{31} - 10^{33}\ m^{-3}$) and short energy confinement time. On the other hand, magnetic confinement uses low plasma density ($10^{18} - 10^{20}\ m^{-3}$) with long energy confinement time. Laser fusion despite being a fantastic technological achievement is less promising as a base-load energy source. Magnetic confinement fusion, on the other hand, holds the current world record of producing 16 MW of fusion power which was achieved at the Joint European Torus (JET) tokamak in 1997. Very recently, Experimental Advanced Superconducting Tokamak (EAST) in China successfully produced and contained the plasma at a temperature of close to 50 million degrees Celsius for an impressive 102 s. This was shortly preceded by Germany's Wendelstein 7-X which heated the fusion fuel to 80 million degrees Celsius and contained it for quarter of a second.

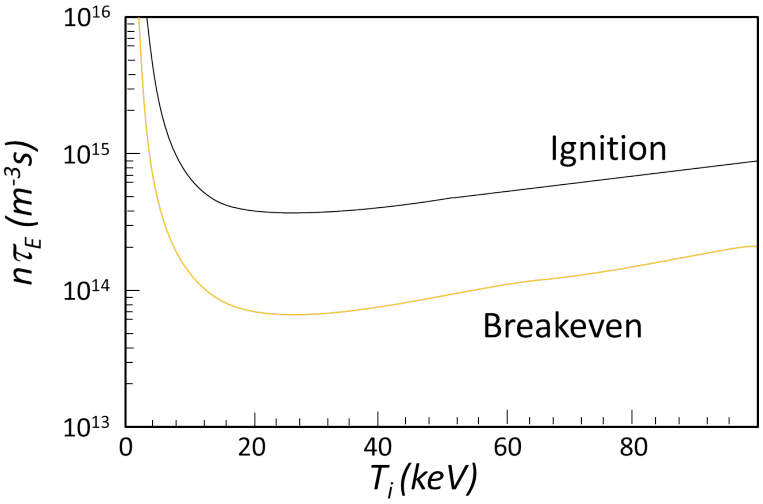


Figure 1.2: *Lawson criterion for DT fusion. Figure adapted from [6]. The criterion states that the product of density and confinement time has to be higher than a value that varies with ion temperature. The lower curve marked "breakeven" refers to the scenario where fusion energy just balances the input energy. The upper curve labeled "ignition" refers to a self-sustaining plasma. The curves have been recomputed using data from [7] and assuming thermal efficiency of 30%.*

1.3 Magnetic confinement

Charged particles, in the presence of magnetic fields are acted upon by the *Lorentz* force which causes them to gyrate in the so-called *Larmor* orbits. Charged particles are, however, free to move parallel to the magnetic field. The gyro-frequency for the traversed helical paths, is given by

$$\omega_c = \frac{qB}{m}, \tag{1.3}$$

where, q is the charge of the particle, m is the particle mass and B is the magnetic field strength. This then corresponds to a gyro-radius of

$$\rho = \frac{v_\perp}{\omega_c}, \tag{1.4}$$

where, $v_\perp$ is the velocity of the particle perpendicular to the magnetic field. The magnetic field lines require to be helical for offsetting the drifts due to $\boldsymbol{E} \times \boldsymbol{B}$, curvature of $\boldsymbol{B}$ and $\boldsymbol{\nabla B}$ which would lead to a loss of confinement otherwise.The two main type of devices for magnetic confinement of fusion plasmas are: stellarators (such as Wendelstein 7-X) and tokamaks. In stellarators, twisted coils are used for producing helical magnetic field whereas in tokamaks, a helical magnetic field is obtained by a combination of two magnetic fields.

This thesis focuses on magnetically confined plasmas in tokamaks.

1.3.1 Tokamak operation

The *Tokamak*[1] is a fusion reactor concept invented in the 1950s by I.Tamm and A.Sakharov [8]. In Figure 1.3 the principal components of a tokamak are shown. Toroidal field coils placed around the plasma generate the toroidal magnetic field. In addition, a current is driven through the plasma, which generates a poloidal magnetic field perpendicular to the toroidal field, resulting in helically twisted magnetic field lines winding around the torus. The current is obtained by using the plasma as secondary circuit of a transformer, whose primary circuit is formed by the inner poloidal field coils.

[1] A Russian acronym for "toroidal chamber with magnetic coils".

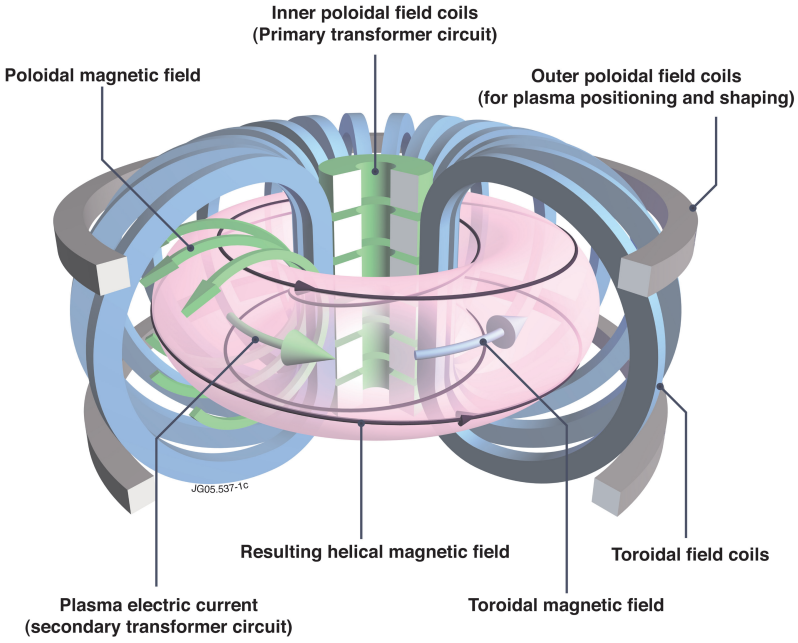


Figure 1.3: *Schematic of a tokamak. The inner poloidal field coils induce a toroidal plasma current and hence provide the poloidal magnetic field. Coupled with the toroidal field, this results in a helical magnetic field. Figure reproduced from EUROfusion [9].*

Further, the outer poloidal field coils are used for position control and shaping of the plasma. Additionally, their vertical magnetic field induces a force that compensates the hoop force due to the plasma kinetic pressure and the $\boldsymbol{j} \times \boldsymbol{B}$ force.

In addition to the magnetic confinement components described in Figure 1.3 there are also several other important parts of a tokamak: the vacuum vessel, the blanket, divertor, heating systems and diagnostics (magnetic systems, spectroscopic instruments etc). The JET tokamak, where the bulk of this work has been carried out, is currently the world's largest operating tokamak. ASDEX Upgrade (AUG), where a portion of this work has also been performed, is a medium sized tokamak. The key engineering parameters for both JET and AUG are given in Table 1.1. The world's largest magnetic confinement fusion device *International Thermonuclear Experimental Reactor* (ITER) which is currently under construction in southern France, is also a tokamak device. ITER is intended to bridge the gap between today's smaller-scale experimental fusion devices and the demonstration fusion reactor (*DEMO*) which is expected to follow ITER. ITER's goal is to produce a ten-fold return on energy ($Q = 10$) and sustain a stable plasma for longer durations. Engineering parameters for ITER are also given in Table 1.1 to allow an ease of comparison with AUG and JET.

Parameter	AUG	JET	ITER
Major radius (R_o)	1.65 (m)	2.96 (m)	6.20 (m)
Minor radius (a)	0.50 (m)	1.25 (m)	2.00 (m)
Plasma volume (V)	13.0 (m^3)	100 (m^3)	840 (m^3)
Plasma current (I_p)	2.00 (MA)	4.80 (MA)	15.0 (MA)
Toroidal magnetic field (B_t)	3.90 (T)	3.45 (T)	11.8 (T)
Plasma heating (P_{heat})	27 (MW)	38 (MW)	50 (MW)

Table 1.1: *Engineering parameters for ASDEX Upgrade (AUG), Joint European Torus (JET) and the ITER tokamaks.*

In order to avoid tritium retention and an erosion of the carbon fiber composite (CFC) divertor target, ITER will operate with beryllium (Be) in the main wall chamber and tungsten (W) in the divertor during its active phase of operation [10][11]. In order to improve predictions for ITER, carbon plasma facing components (PFCs) (hereafter carbon wall or CW) at JET were replaced in 2010 by Be in the main chamber and W in the divertor (hereafter ITER-like wall or ILW) [12]. AUG has operated with an ITER relevant full W wall since approximately 2007[13].

1.3.2 H-mode and edge-localized modes

The high confinement mode (H-mode) [14][15], first discovered at ASDEX in 1982, is a particular plasma regime characterized by enhanced energy and particle confinement time. When the heating power exceeds a certain threshold (P_{LH}), the plasma undergoes a transition from a low confinement (or L-mode) state to H-mode. As illustrated in Figure 1.4, H-mode is characterized by steep gradients in temperature and density in the last ∼2 cm of the plasma radius (plasma edge). This results in an edge transport barrier (ETB) [16] that suppresses energy and particle transport and increases confinement, in contrast to the L-mode where confinement is marred by turbulent transport across the magnetic flux surfaces. The region of the ETB is also called the *pedestal*, as the core profile appears to be elevated on top of the ETB. Plasma performance in the core can be improved further when an internal transport barrier (ITB) is formed [17]. Nevertheless, the plasma is prone to several instabilities that can degrade the overall confinement. Magnetic islands and sawteeth are some of the instabilities that can occur in the core of both L and H mode plasmas. In the edge region, the steep pressure gradient in the ETB gives rise to a magnetohydrodynamics instability known as *edge-localized modes* (ELMs) [19]

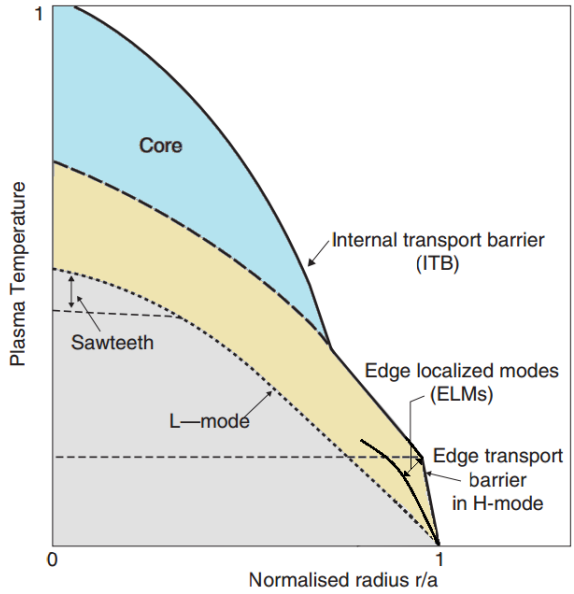


Figure 1.4: *Typical temperature profile observed in L and H mode phases. H-mode lifts the L-mode temperature profile unto the edge transport barrier (ETB). Internal transport barriers are beneficial for the core temperature whereas instabilities like sawteeth are detrimental. Effect of ELMs is also illustrated. The temperature profile collapses as ELMs expel particles and energy from the plasma. Figure adapted from [18]*

[20]. ELMs are short, intense disturbances of the plasma edge which eject energy and particles from the plasma and can cause high transient heat loads on the PFCs. The expulsion of edge plasma brings a reduction in the edge pedestal height (shown in Figure 1.4) which leads to a degradation in energy confinement. Detrimental effects on the PFCs, such as melting, erosion and evaporation, are expected to scale with the size of the tokamak. For medium sized tokamaks ($a \approx 0.5m$), ELMs are not found to cause any damage to the PFCs, however, for JET ($a \approx 1.0m$) melting of the Be divertor surfaces has been observed [21]. For ITER ($a \approx 2.0m$), power flux released by large ELMs will cause an intolerable erosion and heat load on the PFCs [22][23].

Despite this, ELMs are not entirely disadvantageous as the particle exhaust caused by them contributes to the prevention of impurity accumulation in the plasma. Further, ELMs also aid in the removal of helium ash (see (1.1)) from the core which can otherwise suffocate the plasma [24].

H-mode is the reference scenario for ITER's ten fold power multiplication ($Q \sim 10$) inductive operation [25]. This implies that comprehensive physical understanding of ELMs and their control and mitigation is crucial for reaping the benefits of good energy confinement provided by the H-mode. ELM physics and

their control and mitigation are subject of an intense research effort world-wide. This thesis, in line with this effort, focuses on the analysis of ELMs, however, using advanced data analysis.

1.4 Advanced data analysis: Pattern recognition

The last half century witnessed unprecedented advances in the collection, transmission and storage of data. This simultaneously led to the evolution of a new generation of different techniques, methods and algorithms to assist researchers, analysts, decision makers and managers in extracting useful patterns from the rapidly growing volumes of data. These techniques and tools constitute a domain which we here refer to as *advanced data analysis*. Advanced data analysis as shown in Figure 1.5 has evolved from the interaction and cooperation among different sub-fields such as machine learning, knowledge discovery in databases (KDD), statistics and pattern recognition. Advanced data analysis techniques and in particular the domain of *pattern recognition* [26] [27] has transformed a wide variety of industries. Pattern recognition has well-established applications in speech recognition, bioinformatics, remote sensing, biometric recognition, multimedia database retrieval, industrial automation and autonomous navigation [28]. More specifically, Google uses pattern recognition for identifying webspam and e-commerce and technology companies like Braintree use it for stopping credit card fraud. Pattern recognition methods are also brought to use at the Laser Interferometer Gravitational-wave Observatory (LIGO) for identifying gravitational-wave signals from non-Gaussian noise artifacts [29] [30].

Pattern recognition entails the process of discovering patterns and useful knowledge from data. Often considered synonymous with machine learning, it seeks incrementally to understand, adapt and apply these patterns to future cases or data sets. Pattern recognition holds significant potential for improving engineering and control as well as enhancing the physical understanding of fusion plasmas which are faced with many veritable data analysis challenges, such as:

- Large volumes of data are generated in fusion devices. As of 2009 more than

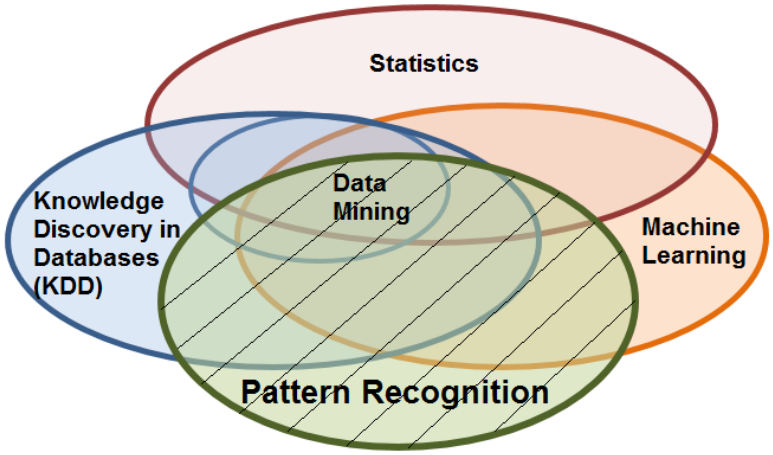


Figure 1.5: *A schematic depicting the relationships between different sub-fields of advanced data analysis.*

10 GB data can be generated per shot at JET [31]. Further, the growth rate of the database roughly follows a Moore's law like doubling every 2 years [32]. Scaling to next step fusion devices indicate that in ITER the volume of data will be orders of magnitude larger.

- The data is often high dimensional i.e described by many physical variables.
- There are complex nonlinear interactions and dependencies between parameters which give rise to data redundancy.
- Unlike high energy physics, where the primary task consists of isolating the products of a specific reaction, fusion plasmas require comprehensive understanding of all data collected in each single discharge.
- There are significant *uncertainties*:
  - Hot fusion plasmas are not very accessible for measurement therefore many parameters are derived indirectly by specialized instrumentation called *plasma diagnostics*. Thus, measurements obtained in fusion experiments are usually hampered by considerable random as well as systematic errors.
  - Several important phenomena in fusion plasmas such as disruptions, turbulence and ELMs are often non-linear and non deterministic. This implies that they behave in a random and unpredictable way due to the complex underlying microscopic physics.

The standard framework to deal with data in the presence of uncertainty is probability theory. This doctoral work starts from the point of view that the fundamental object resulting from a measurement is in fact a probability distribution, with every single measurement providing a sample from this distribution. The complete distribution describing a measured variable potentially contains much more useful information than an average or individual measurements (or collections thereof).

Hence, in this work pattern recognition in spaces of probability distributions is performed. The *Rao geodesic distance* (GD) (Chapter 2) is used as a natural and theoretically well-motivated similarity measure between probability distributions.

1.5 This thesis

This doctoral work is both fundamental and applied. On the one hand pattern recognition methods in geometric spaces of probability distributions are developed and validated. On the other hand, they are applied to the analysis of ELMs. Understanding the underlying physics of ELMs and developing a consistent model for predicting ELM losses is crucial for the success of ITER and next step fusion devices.

1.5.1 Outline

This thesis is structured as follows:
A brief introduction to nuclear fusion by magnetic confinement, H-mode regime and ELMs as well as the motivations for using pattern recognition techniques has been given in this chapter.

Chapter 2 starts by discussing the essential features of the domain of pattern recognition. It distinguishes between *supervised* and *unsupervised* learning and then presents the framework for distance based pattern recognition in non-Euclidean spaces of probability distributions using the mathematical domain of *information geometry* [33]. The theory of pattern recognition methods developed and applied in this work is presented herein.

Chapter 3[2] presents the experimental results of the application of

[2]Chapter 3 is the transcripts of published papers

distance-to-centroid classifier (D2C) and *principal geodesic classifier* (PGC) to multivariate texture discrimination. This serves as a validation of the developed methods and also illustrates the utility and adaptability of the techniques developed in this work to a wide range of applications other than ELMs.

Chapter 4 completes the introduction on H-mode physics and then provides a review of ELM phenomenology and control.

Chapter 5 presents the application of the information visualization methods developed for the visualization of the multi-dimensional and often complex nonlinear data characterizing the operational space of a tokamak onto a 2-D map. The tools are then applied for detecting cluster structure corresponding to type I and type III ELMs in the JET CW dataset and for visualizing the confinement data from the International Tokamak Physics Activity (ITPA) Global H-mode Confinement Database (ITPA database).

Chapter 6[3] and chapter 7 focus on the development of an automatic classification scheme for ELM types with the aim to distinguish ELM classes in a practical, fast and standardized way.

To this end, Chapter 6 presents ELM regime classification (regimes with small ELMs vs type I ELMs) using GD-based conformal predictors based on measurements of global plasma parameters and their error bars from the ITPA database. The classifications are accompanied with estimates of their accuracy and reliability. The estimates of goodness of the predictions increase the knowledge about the accessibility of the various ELMy regimes, while allowing more reliable decisions regarding plasma control.

Chapter 7 presents physical characterization of the JET operational space regarding ELMs by means of discriminant analysis. A parametric classification system for ELM types is presented and applied to the classification of type I and type III ELMs in a set of JET CW plasmas. Further, linear discriminant functions are constructed for determining the boundary between type I and type III ELMy regimes in terms of global plasma parameters. The functions provide an insight into the dependence of the boundary on the plasma and machine conditions and identify the parameters which contribute most to the type I/III boundary. In the second part, a GD-based k-nearest neighbour (kNN) classification scheme is presented to

[3]Chapter 6 is the transcript of a published paper

allow for an effective treatment of the distributions of plasma quantities. A threshold based ELM detection algorithm is developed for the extraction of inter-ELM time intervals (also referred to as waiting times). Waiting times are then modeled with suitable probability distribution functions (PDFs) which are used for GD-based classification of ELMs in JET CW plasmas and a small dataset of AUG plasmas.

In Chapter 8[4] the relationship between ELM energy loss and waiting time for individual ELMs is studied in contrast to examining the average values over a discharge. The analysis is conducted on a set of unseeded JET ILW plasmas, N_2 seeded ILW plasmas and CW plasmas. Further, the impact of slow transport events (STEs)[34][35] on the distribution of ELM durations as well as on the correlation between ELM energy loss and ELM waiting time is examined. Lastly, regression analysis is conducted for determining the regime, in terms of global plasma parameters, which will maximize the correlation between ELM energy loss and waiting time.

Finally, conclusions and an outlook towards future research topics is presented in Chapter 9.

1.5.2 List of publications

In this section a list is given of the publications acquired during this doctoral work. Also provided is a list of oral and poster presentations made during the course of this PhD.

Journal publications

- **A.Shabbir**, G.Hornung, G.Verdoolaege and JET contributors,"A classification scheme for edge-localized modes based on their probability distributions," *Review of Scientific Instruments*, vol.87, issue 11, no.11D404, 2016.
- G.Verdoolaege, **A.Shabbir** and G.Hornung,"Robust analysis of trends in noisy diagnostic data," *Review of Scientific Instruments*, vol.87, issue 11, no.11D422, 2016.
- G.Hornung, **A.Shabbir** and G.Verdoolaege, "A first approach towards Bayesian estimation of turbulent plasma properties from reflectometry," *Fusion Science and Technology*, vol.69, no.3, pp.586-594, 2016.
- **A.Shabbir**, G.Verdoolaege, J.Vega and A.Murari, "ELM classification by conformal prediction on an information manifold," *IEEE Transactions on Plasma Science*, vol.43, no.12, pp.4190-4199, 2015.

[4]Chapter 8 is in the format of a transcript which is due for submission

- **A.Shabbir**, G.Verdoolaege, O.J.W.F.Kardaun, J.-M.Noterdaeme and JET-EFDA Contributors, "Visualization of the operational space of edge-localized modes through low-dimensional embedding of probability distributions," *Review of Scientific Instruments*, vol.85, issue 11, no.11E819, 2014.

To be published:

- **A.Shabbir**, G.Verdoolaege, G.Hornung, O.J.W.F Kardaun, H.Zohm and JET contributors, "Correlation analysis for energy losses, waiting times and durations of type I edge-localized modes in the Joint European Torus," status: under review.

- **A.Shabbir**, G.Hornung, J.-M.Noterdaeme, G.Verdoolaege and JET contributors,"Physical characterization of the JET operational space regarding ELMs by means of discriminant analysis," status: to be submitted.

Conference publications

- **A.Shabbir**, G.Verdoolaege, G.Hornung, O.J.W.F Kardaun, H.Zohm and JET contributors, "Analysis of the energy losses and waiting times for individual ELMs in the Joint European Torus," in *Proc.* 43^{rd} *EPS Conference on Plasma Physics*, Leuven, Belgium, July, 2016.

- **A.Shabbir** and G.Verdoolaege, "Multivariate texture discrimination using a principal geodesic classifier," in *Proc. IEEE International Conference on Image Processing*, pp. 3550-3554, Quebec City, Canada, September, 2015.

- G.Verdoolaege and **A.Shabbir**, "Color texture texture discrimination using the principal geodesic distance on a multivariate generalized Gaussian manifold," in *Geometric Science of Information, ser. Lecture Notes in Computer Science*, vol.9389, pp. 379-386, 2015.

- **A.Shabbir**, G.Verdoolaege, O.J.W.F.Kardaun, A.J. Webster, R.O. Dendy, J.-M.Noterdaeme, and JET-EFDA Contributors, "Discrimination and visualization of ELM types based on a probabilistic description of inter-ELM waiting times," in *Proc.* 41^{st} *EPS Conference on Plasma Physics*, P1.001, Berlin, Germany, June, 2014.

- **A.Shabbir**, G.Verdoolaege and G.Van Oost, "Multivariate texture discrimination based on geodesics to class centroids on a generalized Gaussian manifold." in *Geometric Science of Information, ser. Lecture Notes in Computer Science* , vol. 8085, pp. 853-860, 2013.

- **A.Shabbir**, G. Verdoolaege, G.Van Oost, J.-M.Noterdaeme, and JET-EFDA Contributors, "Visualization of tokamak operational spaces through the projection of data probability distributions," in *Proc.* 40^{th} *EPS Conference on Plasma Physics*, P2.132, Espoo, Finland, July, 2013.

Oral presentations

- **A.Shabbir**, G.Hornung and G.Verdoolaege "Identification of edge-localized instabilities in nuclear fusion plasmas using pattern recognition techniques," *General Scientific Meeting 2016 of the Belgian Physical Society*, Ghent, Belgium, May, 2016.

- **A.Shabbir** and G.Verdoolaege, "Multivariate texture discrimination using a principal geodesic classifier," in *IEEE International Conference on Image Processing*, Quebec City, Canada, September, 2015.

- **A.Shabbir**, G.Verdoolaege, G.Hornung, O.J.W.F Kardaun, H.Zohm and JET contributors, "Analysis of the dependence between inter-ELM time intervals and energy losses," in 1^{st}*IAEA technical meeting on Fusion Data Processing, Validation and Analysis*, Nice, France, June, 2015.

- **A.Shabbir**, G.Verdoolaege, A.J. Webster, R.O. Dendy, J.-M. Noterdaeme and JET-EFDA Contributors, "Discrimination and visualization of edge-localized modes on an information manifold," in 8^{th} *Workshop on Fusion Data Processing, Validation and Analysis*, Ghent, Belgium, November, 2013.

- **A.Shabbir**, G.Verdoolaege and G. van Oost, "Pattern recognition for real time control of fusion plasma," 13^{th} *FEA PhD symposium*, Ghent, Belgium, December, 2012.

Poster presentations

- **A.Shabbir**, G.Hornung, G.Verdoolaege and JET contributors,"K-nearest neighbor classification on an information manifold and its application to edge-localized modes in fusion plasmas," 36^{th} *International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering (MaxEnt 2016)*, Ghent, Belgium, July, 2016.

- **A.Shabbir**, G.Verdoolaege, G.Hornung, O.J.W.F Kardaun, H.Zohm and JET contributors, "Analysis of the energy losses and waiting times for individual ELMs in the Joint European Torus," 43^{rd} *EPS Conference on Plasma Physics*, Leuven, Belgium, July, 2016.

- **A.Shabbir** and G.Verdoolaege, "Machine learning in probability spaces for optimizing control in fusion plasmas," 29^{th} *Machine Learning Summer School*, Kyoto, Japan, September, 2015.

- **A.Shabbir**, G.Hornung, G.Verdoolaege and JET contributors, "Discriminant analysis for edge-localized plasma instabilities," 1^{st} *EPS Conference on Plasma Diagnostics*, Frascati, Italy, April, 2015.

- G.Hornung, **A.Shabbir** and G.Verdoolaege, "Bayesian inference of plasma turbulence properties from reflectometry measurements," 1^{st} *EPS Conference on Plasma Diagnostics*, Frascati, Italy, April, 2015.

- **A.Shabbir** and G.Verdoolaege, "Pattern recognition in probability spaces for the analysis of edge-localized plasma instabilities," 4^{th} *FuseNET PhD Event*, Lisbon, Portugal, November, 2014.

- **A.Shabbir**, G.Verdoolaege, O.J.W.F.Kardaun, J.-M.Noterdaeme and JET-EFDA Contributors, "Visualization of the operational space of edge-localized modes through low-dimensional embedding of probability distributions," 20^{th} *Topical Conference on High-Temperature Plasma Diagnostics*, Atlanta, GA, USA, June, 2014.

- **A.Shabbir**, G.Verdoolaege, O.J.W.F.Kardaun, A.J. Webster, R.O. Dendy, J.-M.Noterdaeme, and JET-EFDA Contributors, "Discrimination and visualization of ELM types based on a probabilistic description of inter-ELM waiting times," 41^{st} *EPS Conference on Plasma Physics*, Berlin, Germany, June, 2014.

- **A.Shabbir**, G.Verdoolaege and G.Van Oost, "Multivariate texture discrimination based on geodesics to class centroids on a generalized Gaussian manifold." *Geometric Science of Information Conference*, Paris, France, August, 2013.

- **A.Shabbir**, G.Verdoolaege, G.Van Oost and J.-M. Noterdaeme, "Probabilistic pattern recognition for improving plasma control and physical understanding," 50^{th} *Culham Plasma Physics Summer School* Culham, Oxfordshire, UK, July, 2013.

- **A.Shabbir**, G. Verdoolaege, G.Van Oost, J.-M.Noterdaeme, and JET-EFDA Contributors, "Visualization of tokamak operational spaces through the projection of data probability distributions," 40^{th} *EPS Conference on Plasma Physics*, Espoo, Finland, July, 2013.

- **A.Shabbir**, G.Verdoolaege, G.Van Oost and J.-M. Noterdaeme, "Finding meaning in randomness: pattern recognition in probability spaces for effective plasma control and enhanced physical understanding", *3rd FuseNET PhD Event*, York, UK, June, 2013.

Chapter 2

Pattern recognition in spaces of probability distributions

This chapter covers the theoretical foundations and background of pattern recognition methods which are important for a profound understanding of the contents of this thesis. The framework of pattern recognition in non-Euclidean spaces of probability distributions is laid out and followed by a description of pattern recognition methods which are adapted and developed in this work.

2.1 Fundamentals of pattern recognition

A *pattern* is essentially a commonality among the multiple instances of an entity. More formally, S. Watanabe [36] defines a pattern as opposite of chaos and an entity vaguely defined that could be given a name. Given a pattern, its recognition comes naturally to human beings. Humans can recognize faces without conscientiously processing varying illuminations, facial rotations, facial expressions etc. Humans can with relative ease recognize the sound of a human voice from that of a violin; the aroma of a rose, from that of garlic and a numeral '4' from a numeral '8'. Humans essentially learn from experience. Pattern recognition is the study of how machines (such as computer programs) can learn from an observation of the environment, distinguish patterns of interest from their background, and make sound and reasonable decisions about the categories of the patterns. Pattern recognition also includes preprocessing procedures to normalize data, to deal with invariants and to define proper features and distance measures.

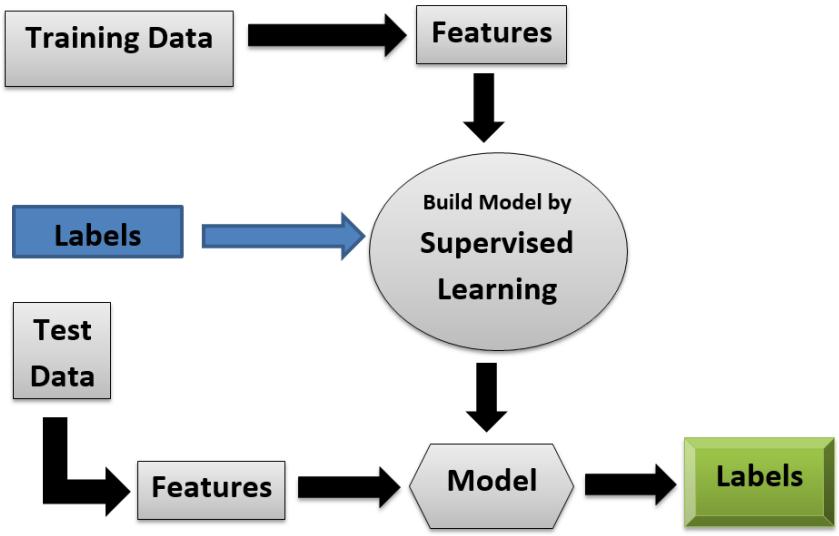


Figure 2.1: *Workflow of supervised learning.*

Physical systems are empirically and scientifically investigated by the following steps [37]:

- Data collection
- Preprocessing
- Model building (learning)
- Prediction or knowledge discovery

Depending on data collection, learning problems can be broadly divided into two categories: *supervised* and *unsupervised*. In the first case, data is observed with the outcome measurement which is also referred to as the *class label*. Preprocessing and model building extracts information from the data for characterizing the underlying process and is guided by the outcome variable. The workflow of supervised learning is given in Figure 2.1. In the case of unsupervised learning, illustrated in Figure 2.2, the observed data does not have an associated outcome variable. Patterns are usually discovered by grouping the observations in the model building stage and this aids in the understanding of the underlying physical process that generated the data. Preprocessed data that is used for building the model (learning) is known as *training* data whereas the data that is used for assessing the strength and predictive capability of the model is called *test* data. Each preprocessed data observation is represented by N features or (physical) quantities and is considered a point in a N dimensional space. As shown in Figure 2.3, supervised learning

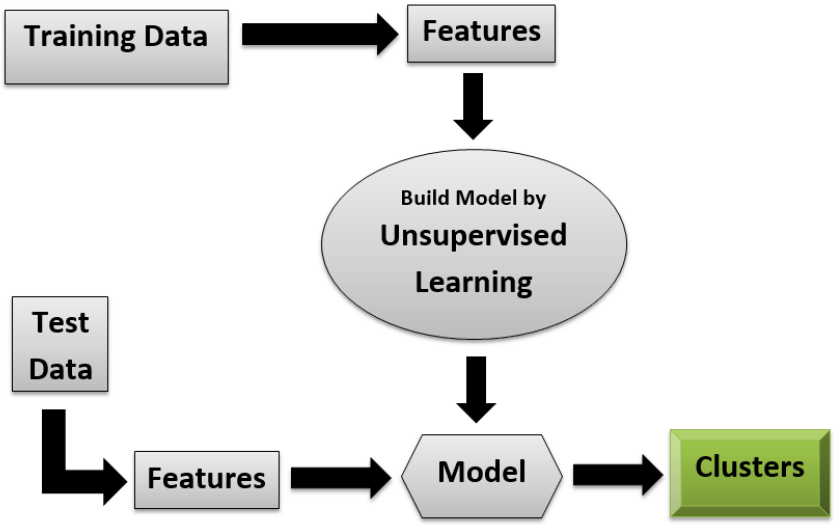


Figure 2.2: *Workflow of unsupervised learning.*

can manifest itself as classification or regression methods whereas unsupervised learning corresponds to clustering applications. Preprocessing methods include correlation analysis, dimensionality reduction, feature extraction, feature selection, normalization and standardization amongst others. In this dissertation, we focus primarily on supervised learning methods (classification, regression), dimensionality reduction and correlation analysis.

- *Correlation analysis:* Correlation measures the statistical relationship involving dependence between two sets of data or random variables. An existence of correlation nullifies probabilistic independence and tends to indicate a predictive relationship that can be exploited in practice. Correlation analysis is widely used in medicine and social sciences for examining relations such as those between education and income, unemployment and crime, maternal age and infant mortality etc.

- *Dimensionality reduction:* The dimension of the data is the number of variables that are measured or that characterize each observation. Mathematically, dimensionality reduction entails finding for a N-dimensional random variable $p = (p_1, ...p_N)$ a lower dimensional representation $q = (q_1, ...q_k)$ with $k \leq N$, that captures the content in the original data, according to some criterion. Advances in data collection and storage capabilities during the past decades have led to an exponential growth of high dimensional data in most sciences including nuclear fusion [31][32]. In most cases, dimensionality

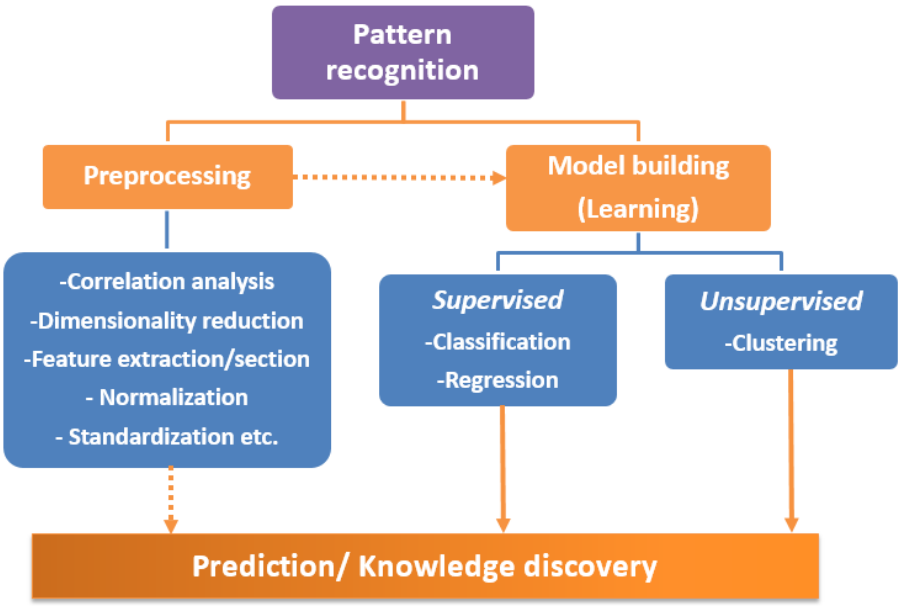


Figure 2.3: *Schematic of pattern recognition.*

reduction methods contribute to improving the speed and accuracy of learning algorithms, economizing data storage requirements and enhancing the understanding of underlying phenomena of interest by discarding *irrelevant* and *redundant* variables.

- *Supervised learning:*

  – *Classification:* It is the identification of the category to which a new observation (or instance) belongs based on the training data containing observations whose category membership is known. Examples include assigning a given email into "spam" or "non-spam" folders, assigning diagnosis to a given patient based on diet and clinical measurements, prediction of onset of plasma disruption [38] [39] and determination of ELM type and disruption type in tokamak plasmas [40][41].

  – *Regression*: As a workhorse from statistics, regression analysis assigns a real valued outcome to a new observation based on the model learned from the labeled training data. Regression analysis is used extensively in fusion plasma physics for fitting deterministic relations reflecting physical dependencies between plasma variables. Further, scaling laws such as those for energy confinement time and power threshold for the L to H mode transition have been derived from multi-machine databases [42].

More recently, advanced regression methods such as symbolic regression [43] and geodesic least squares regression [44] have been applied in fusion for circumventing the limitations of ordinary least squares regression.

2.2 Geometric probabilistic framework

As introduced in Chapter 1, measurements in fusion experiments can be affected by considerable uncertainties, both systematic and stochastic. Probability theory in the presence of uncertainty and non-deterministic phenomena provides a natural description of the raw data. Each measurement x is regarded as a sample from an underlying probability distribution of the measurement characterized by its probability density function (PDF), $p(x|\theta)$. Measurements described by the distribution parameters, θ, may contain significantly more information than a measurement expressed as a value and an associated error bar. In this work, we apply pattern recognition methods directly in a probabilistic data space i.e. a space of probability distributions. Since pattern recognition essentially relies on quantitative assessment of the proximity of data points, use of a well-suited similarity (distance) measure for quantitatively comparing $PDFs$ is necessitated.

Similarity measures

Similarity measures that satisfy the fundamental definitions of a *metric* as identified widely in the mathematical literature [45] are called distance measures while other non-metric similarity measures are occasionally called *divergence*. The definition of a distance measure includes the following requirements:

1. Non-negativity: $dist(A, B) \geq 0$

2. Identity of indiscernibles: $dist(A, B) = 0$ iff $A = B$

3. Symmetry: $dist(A, B) = dist(B, A)$

4. Triangle inequality: $dist(A, B) \leq (dist(A, C) + dist(B, C))$

2.2.1 Euclidean distance

In the n-dimensional Euclidean space R^n, the shortest distance between two points is given by the Euclidean distance which is defined as

$$d_{Euc}(P,Q) = \sqrt{\sum_{i}^{n}(Q_i - P_i)^2} \tag{2.1}$$

Euclidean geometry is frequently used in various domains as well as in everyday life. However, it is a poor measure for assessing the similarity of two or more PDFs as it does not treat the intrinsic nature of probability distributions properly.

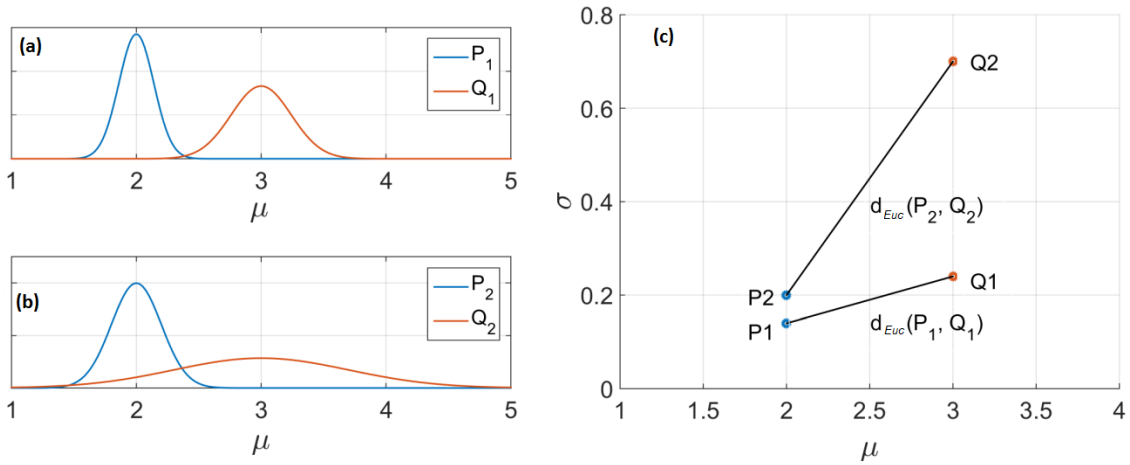


Figure 2.4: *Illustration of the Euclidean distance between probability distributions.*

This is illustrated in Figure 2.4. We consider two Gaussian PDFs $P_1(x|2, 0.15)$ (i.e. $\mu = 2$, $\sigma = 0.15$) and $Q_1(x|3, 0.25)$ in Figure 2.4(a) and two Gaussians $P_2(x|2, 0.2)$ and $Q_2(x|3, 0.7)$ in Figure 2.4(b). P_2 and Q_2 have the same respective means as P_1 and Q_1 but larger standard deviations. It can be seen from Figure 2.4 that PDFs P_2 and Q_2 have a larger overlap and appear more similar in comparison with P_1 and Q_1. However, the Euclidean space representation in Figure 2.4(c) suggests the opposite as the Euclidean (straight line) distance between P_2 and Q_2 is larger than the distance between P_1 and Q_1. This illustrates the inadequacy of the Euclidean distance in capturing the real similarity and consequently the physical proximity of PDFs.

2.2.2 Kullback-Leibler divergence

The *Kullback-Leibler divergence* (KLD) between continuous distributions P_1 and P_2 with respective PDFs $p_1(x)$ and $p_2(x)$ is defined as,

$$KL(P_1, P_2) = \mathbb{E}_{P_1}[\ln L] = \int_{-\infty}^{\infty} p_1(x) \ln \frac{p_1(x)}{p_2(x)} \tag{2.2}$$

where $L = \frac{p_1(x)}{p_2(x)}$ is the likelihood ratio. Therefore,

$$KL(P_1, P_2) = -\int_{-\infty}^{\infty} (p_1(x)\ \ln\ p_2(x)) + \int_{-\infty}^{\infty} (p_1(x)\ \ln\ p_1(x)) = H(P_1, P_2) - H(P_1), \tag{2.3}$$

where $H(P_1)$ is the *entropy* of P_1 and $H(P_1, P_2)$ is the *cross-entropy* of P_1 and P_2.

KLD, also known as information gain or relative entropy, is a popular similarity measure for probability distributions. However, it exhibits certain disadvantages. Firstly, it falls short of being a genuine distance measure as it does not satisfy the symmetry condition and the triangle inequality presented in section 2.2 as fundamental definitions of a distance measure. The triangle inequality, which the KLD does not obey, is a useful property for reducing the computational demands in various applications such as image retrieval [46][47]. Secondly, KLD computation requires calculating a multidimensional integral over the data space. Therefore, closed form expressions for the KLD are difficult to find and numerical estimation inevitably increases the computational load significantly. Thirdly, it has recently been demonstrated that compared with KLD, the *Rao geodesic distance* (GD) presented in the next section is a more accurate similarity measure between probability distributions [48] [49][50].

2.2.3 The Rao geodesic distance

In differential geometry the notion of a *geodesic* is a generalization of the notion of a *straight line* to curved spaces. The term finds its roots in Geodesy which is the science of measuring the size and shape of the Earth. The mathematical field of information geometry [33] [51] allows a probability density family (likelihoods, probabilistic models) to be interpreted as a Riemannian differentiable manifold. A point on the manifold corresponds to a specific probability density function (PDF)

within the family and the family parameters provide a coordinate system on the manifold [48]. Cramér[52] and Rao[53] noted that the Fisher information provides a metric tensor (Fisher-Rao) on the manifold of probability distributions. The Fisher-Rao metric is a unique intrinsic metric on such a manifold and is invariant under some basic probabilistic transformations [54]. For a probability model $p(x|\theta)$ describing a vector x, labeled by an N-dimensional vector θ, the components of the Fisher information matrix $g_{\mu\nu}$ are defined as

$$g_{\mu\nu}(\theta) = -\mathbb{E}\,[\frac{\partial^2}{\partial\theta^\mu \partial\theta^\nu}\,\ln\,p(x|\theta)] \quad , \quad \mu,\nu = 1...N\,, \tag{2.4}$$

where $\mathbb{E}$ signifies expectation with respect to the data vector x. The Fisher-Rao metric paves the way for the calculation of geodesics and the GD between two points (probability distributions) on the manifold [55]. The geodesics between probability distributions have a property of length minimization on the manifold and the ensuing GD is a natural and intrinsic distance measure between probability distributions. This is amplified further through an illustrative example in the next section.

2.2.3.1 Univariate Gaussian distribution

The univariate Gaussian distribution, parameterized by its mean μ and standard deviation σ is defined through the following PDF:

$$f(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2)}\right]. \tag{2.5}$$

The Fisher-Rao metric can be given via the quadratic line element [56]

$$ds^2 = \frac{1}{\sigma^2}d\mu^2 + \frac{2}{\sigma^2}d\sigma^2. \tag{2.6}$$

A closed-form solution exists for the GD between two univariate Gaussian distributions $p_1(x|\mu_1,\sigma_1)$ and $p_2(x|\mu_2,\sigma_2)$, parametrized by their means μ_1 and μ_2 and standard deviations σ_1 and σ_2 and is given as,

$$GD(p_1,p_2) = \sqrt{2}\,\ln\frac{1+\delta}{1-\delta} = 2\sqrt{2}\tanh^{-1}\delta\,, \tag{2.7}$$

$$\delta = \left[\frac{(\mu_1-\mu_2)^2 + 2(\sigma_1-\sigma_2)^2}{(\mu_1-\mu_2)^2 + 2(\sigma_1+\sigma_2)^2}\right]^{1/2}.$$

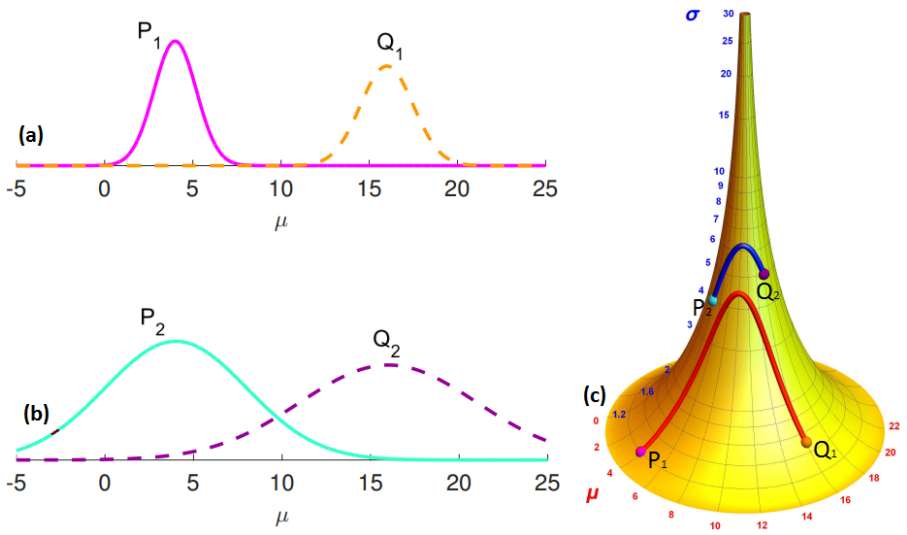


Figure 2.5: *(a) and (b) respectively illustrate the univariate Gaussian PDFs* $P_1(x|4, 1.2)$, $Q_1(x|16, 1.5)$ *and* $P_2(x|4, 4.0)$, $Q_2(x|16, 5.0)$. P_2 *and* Q_2 *have the same mean as* P_1 *and* Q_1 *but a larger standard deviation. In (c), the pseudo-sphere is presented as a model for a univariate Gaussian manifold. The distributions in (a) and (b) and the geodesics between them have been mapped on the surface of the pseudosphere in (c). It can be seen that the distributions* P_2 *and* Q_2 *which have a considerably larger overlap than distributions* P_1 *and* Q_1 *are connected by a visibly shorter geodesic on the manifold in (c). Figure has been adapted from [44].*

	Probability distributions		
Distance measure	P_1 and Q_1	P_2 and Q_2	More similar distributions
GD	5.3	2.4	P_2 and Q_2
Euclidean	12	12.04	P_1 and Q_1

Table 2.1: *GD and Euclidean distance between the Gaussian distributions* P_1 *and* Q_1 *and* P_2 *and* Q_2 *presented in Figure 2.5*

In contrast to the frequently used Euclidean distance measure, the GD respects the intrinsic geometry of the probability distributions. This is illustrated in Figure 2.5. We consider two Gaussian PDFs $P_1(x|4, 1.2)$ (i.e. $\mu = 4$, $\sigma = 1.2$) and $Q_1(x|16, 1.5)$ in Figure 2.5(a) and two Gaussian PDFs $P_2(x|4, 4.0)$ and $Q_2(x|16, 5.0)$ in Figure 2.5(b). P_2 and Q_2 have the same respective means as P_1 and Q_1 but larger standard deviations. GD between distributions has been computed using (2.7) and the results have been listed in Table 2.1. Similarly the Euclidean distance has been computed using (2.4) and the results are presented in Table 2.1. The results presented in Table 2.1 show that the GD between P_2 and Q_2 is significantly higher than the GD between P_1 and Q_1, suggesting higher proximity between P_2 and Q_2. This is in agreement with Figure 2.5 which reveals a significantly higher overlap between P_2 and Q_2 in comparison with P_1 and Q_1. On the other hand, the Euclidean

distance inaccurately predicts P_1 and Q_1 as more similar than P_2 and Q_2 as it does not treat the intrinsically non-Euclidean nature of probability distributions in an effective manner.

An explanatory visualization of the two-dimensional surface of univariate Gaussian PDFs, approximated by the pseudosphere (tractoid) is presented in Figure 2.5(c) [44]. Each point on the surface is a Gaussian PDF parametrized by its mean μ and the standard deviation σ. The meridians represent lines of constant mean, while the latitudes (circles) have a constant standard deviation. Despite being an imperfect model (invalid for $\sigma < 1$) for representing the true geometry of Gaussian distributions, Figure 2.5(c) nevertheless provides an intuitive visualization of the geodesics between the points corresponding to the distributions P_1 and Q_1 and P_2 and Q_2. It can be readily observed from Figure 2.5(c) that the distance between P_2 and Q_2 is indeed shorter than that between P_1 and Q_1.

2.2.3.2 Multivariate generalized Gaussian distribution

The *Multivariate generalized Gaussian distribution*(MGGD), also sometimes called the multivariate exponential power distribution, is a particular case of the multivariate Kotz-type distribution. MGGDs have been widely used in image processing applications. The inclusion of Gaussian and Laplacian distributions as special cases render MGGDs useful for capturing the statistical properties of images or image features. In particular, the distribution of wavelet coefficients has been shown to be effectively modeled by the GGDs [57][58]. This characteristic has been exploited for several applications including content based image retrieval [59], image denoising [60], texture classification [61] and confinement regime identification and disruption prediction in fusion plasmas [62]. In this work, with a view to modeling wavelet detail coefficients (Chapter 3) we consider only zero-mean distributions. As a first step in defining MGGD, we present the PDF for a univariate zero-mean GGD,

$$f(x|\alpha,\beta) = \frac{\beta}{2\alpha\Gamma[1/\beta]} \exp\Big[-(|x|/\alpha)^{\beta}\Big], \tag{2.8}$$

where Γ denotes the Gamma function, α is a scale parameter and $\beta(\beta > 0)$ is the shape parameter. As shown in Figure 2.6(b), α resembles the variance and determines the 'width' of the PDF, while β controls the fall-off rate or the

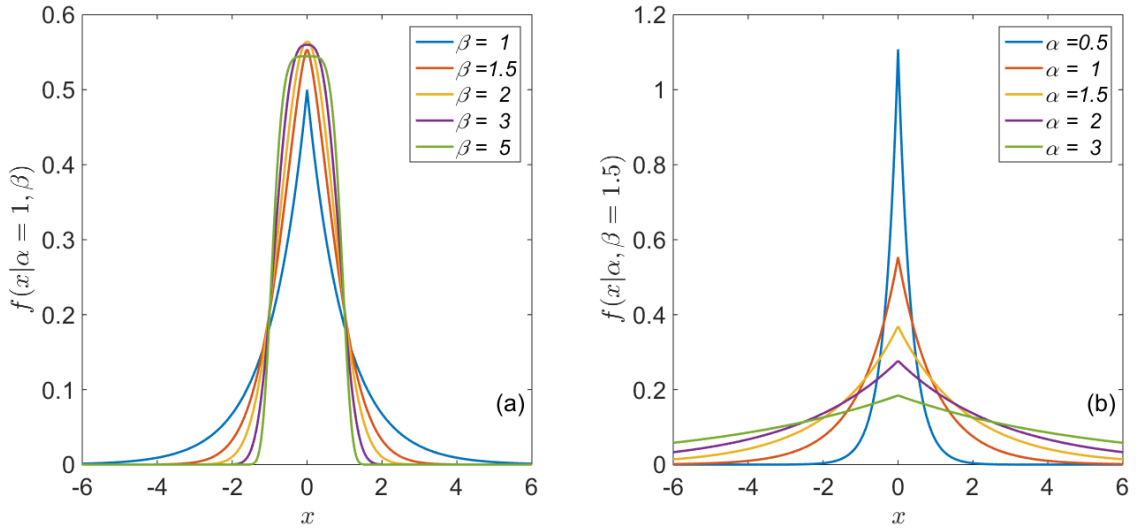


Figure 2.6: *PDF for univariate zero-mean generalized Gaussian distribution. (a). Effect of the change in shape parameter, β, on the distribution. (b). Effect of the change in scale parameter α on the distribution.*

'peakedness' of the distribution. It can be noted from (2.8) and Figure 2.6(a) that $\beta = 2$ results in the Gaussian distribution and $\beta = 1$ yields the Laplacian PDF. We proceed to the multivariate generalized Gaussian distribution, defined in [63] [48] as:

$$f(x|\Sigma, \beta) = \frac{\Gamma(\frac{m}{2})}{\pi^{\frac{m}{2}}\Gamma(\frac{m}{2\beta})2^{\frac{m}{2\beta}}} \frac{\beta}{|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}\Big[X(\Sigma)^{-1}X\Big]^{\beta}\right]. \tag{2.9}$$

Here, m is the dimensionality of the probability space and Σ is the dispersion matrix. The distribution reduces to a multivariate Gaussian case for $\beta = 1$ and to a multivariate Laplace case for $\beta = 0.5$.

Expressions for the metric and the geodesic equations on the manifold of zero-mean MGGDs have been obtained in [55]. For fixed shape parameter β, the GD between MGGDs denoted by (β, Σ_1) and (β, Σ_2) is given in [55][48] as:

$$GD(\beta, \Sigma_1|\beta, \Sigma_2) = \left[\Big(3b_h - \frac{1}{4}\Big)\sum_i (r_2^i)^2 + 2\Big(b_h - \frac{1}{4}\Big)\sum_{i<j} r_2^i r_2^j\right]^{1/2}, \tag{2.10}$$

with $r_2^i \equiv \ln \lambda_2^i$ and λ_2^i, $i = 1, ..., m$, the m eigenvalues of $\Sigma_1^{-1}\Sigma_2$. In addition, b_h is defined by

$$b_h \equiv \frac{1}{4}\frac{m + 2\beta}{m + 2},$$

With variable shape parameter the geodesic equations are more difficult to solve and a closed form for the GD has thus not been obtained. We use a linear approximation

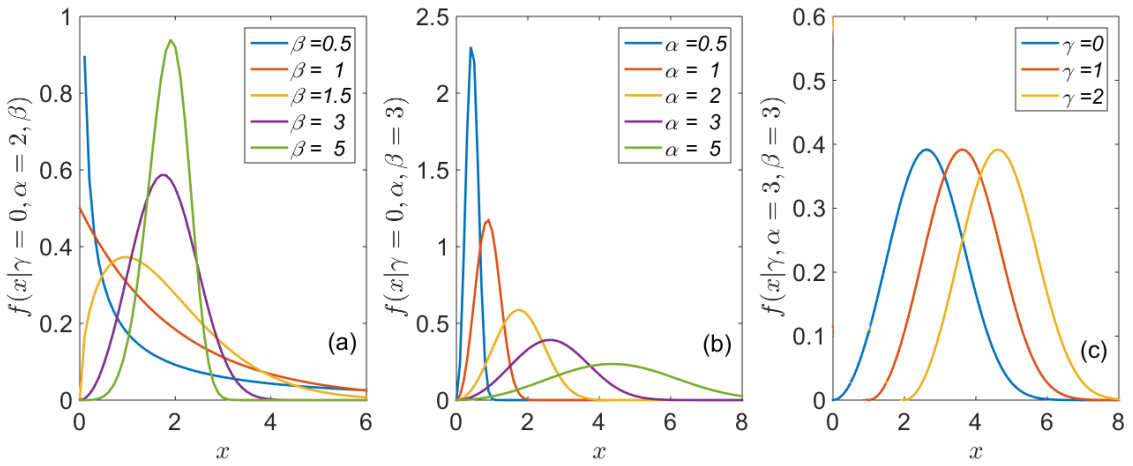


Figure 2.7: *PDF for 3-parameter Weibull distribution. (a). Effect of the change in shape parameter, β, on the distribution. (b). Effect of the change in scale parameter α on the distribution. (c). Effect of the change in location parameter γ on the distribution.*

to the geodesic coordinate functions, to render the calculations computationally more feasible, see [55].

2.2.3.3 Weibull distribution

The Weibull distribution has a wide range of applications and has been frequently used in survival analysis, material sciences, reliability engineering and extreme event modeling.

The 3-parameter Weibull

The 3-parameter (3P) Weibull PDF is given as

$$f(x|\beta,\alpha,\gamma) = \frac{\beta}{\alpha}\left(\frac{x-\gamma}{\alpha}\right)^{\beta-1} \exp\left[-\left(\frac{x-\gamma}{\alpha}\right)^{\beta}\right], \tag{2.11}$$

where, $f(x) \geq 0$, $x \geq 0$ *or* $x = \gamma$, $\beta\,(scale\ parameter) > 0$, $\alpha\,(shape\ parameter) > 0$ and $-\infty < \gamma\ (location\ parameter) < \infty$. As shown in Figure 2.7(a) varying values of the shape parameter β has a marked effect on the behavior of the distribution. For $\beta = 1$, the PDF of the 3P Weibull distribution reduces to that of the 2-parameter (2P) exponential distribution given as:

$$f(x|\alpha,\gamma) = \frac{1}{\alpha} \exp\left[-\left(\frac{x-\gamma}{\alpha}\right)\right]. \tag{2.12}$$

The exponential distribution has the key property of being memoryless. Further, it can be observed from Figure 2.7(c) that increasing the value of the shape parameter α while holding β constant has the effect of stretching out the distribution. Lastly, the effect of the location parameter γ on the PDF is shown in Figure 2.7 (c). It can be noted that varying γ has the effect of sliding the distribution to the right (if $\gamma > 0$) or to the left (if $\gamma < 0$).

The 2-parameter Weibull

The 2-parameter (2P) Weibull PDF is obtained by setting $\gamma = 0$ in (2.11) and is given as:

$$f(x|\beta,\alpha) = \frac{\beta}{\alpha}\left(\frac{x}{\alpha}\right)^{\beta-1} \exp\left[-\left(\frac{x}{\alpha}\right)^{\beta}\right]. \tag{2.13}$$

A closed form solution exists for the GD between two 2P Weibull distributions on a Weibull manifold. For the derivation of the Fisher information metric and the expression for the GD between 2P Weibull distributions refer to [64].

2.3 Dimensionality reduction

Dimensionality reduction, as already introduced, is the transformation of high-dimensional data into a meaningful representation of reduced dimensionality. It is well-aligned with the spirit of Occam's razor: one should not increase, beyond what is necessary, the number of entities required to explain anything. Building on this, ideally, the reduced representation should correspond to the intrinsic dimensionality[65] of the data which is the minimum number of parameters or features needed to account for the observed properties of the data.

Dimensionality reduction in addition to offering a number of attractive advantages is well-motivated for several reasons:

- *Curse of dimensionality*: This term coined by Bellman in 1961 [66] refers to the problems arising due to an exponential increase in the volume of the region of the data space associated with the addition of extra dimensions. A few of these repercussions are:
  - A small increase in dimensionality generally leads to an exponential increase in the quantity of data required for sustaining the same level

of performance for classification, regression etc.

  - *Empty space phenomenon*: High-dimensional spaces are inherently sparse [67]. For instance, for a one-dimensional (1D) standard normal distribution, 70% of the mass is contained in a sphere of radius one standard deviation whereas for a 10-dimensional (10D) standard normal distribution that same (hyper)sphere contains only 0.02% of the mass. For containing 70% mass, a radius of more than 3 standard deviations has to be considered. This implies that in high-dimensional distributions the tails are much more crucial than in 1D ones.

- Original high dimensional data representation may be redundant as:
  - Some parameters might be correlated with each other either through linear combinations or other functional dependences.
  - Some parameters will have a variation smaller than the measurement noise and will thus be *irrelevant*

- *Peaking phenomenon*: For a given size of the data set, the classifier accuracy increases with an increase in the number of parameters (dimensions), peaks to an optimum value and then starts decreasing, with a further addition of parameters. S.Ruadys *et al.* [68] suggested that as the complexity of a classifier increases, the ratio of data size to dimensionality should also be increased in order to avoid the peaking phenomenon.

Dimensionality reduction leads to an improvement in the speed and accuracy of the learning algorithms while yielding more tractable and understandable models. In [69], a feature selection based on genetic algorithms (dimensionality reduction) is performed for improving the prediction capability of the Advanced Predictor of Disruption (APODIS) at JET. The dimensionality reduction not only improves the success rates of APODIS but also extends the interval before the disruption in which reliable predictions are achieved. Similarly, in [70] feature selection (dimensionality reduction) shortlists plasma current (I_p) and mode locked amplitude (ML) along with either plasma internal inductance (LI) or radiated power (P_{rad}) as parameters which give the lowest false alarm rate and a prediction model based on these parameters reduces the computational time for disruption prediction.

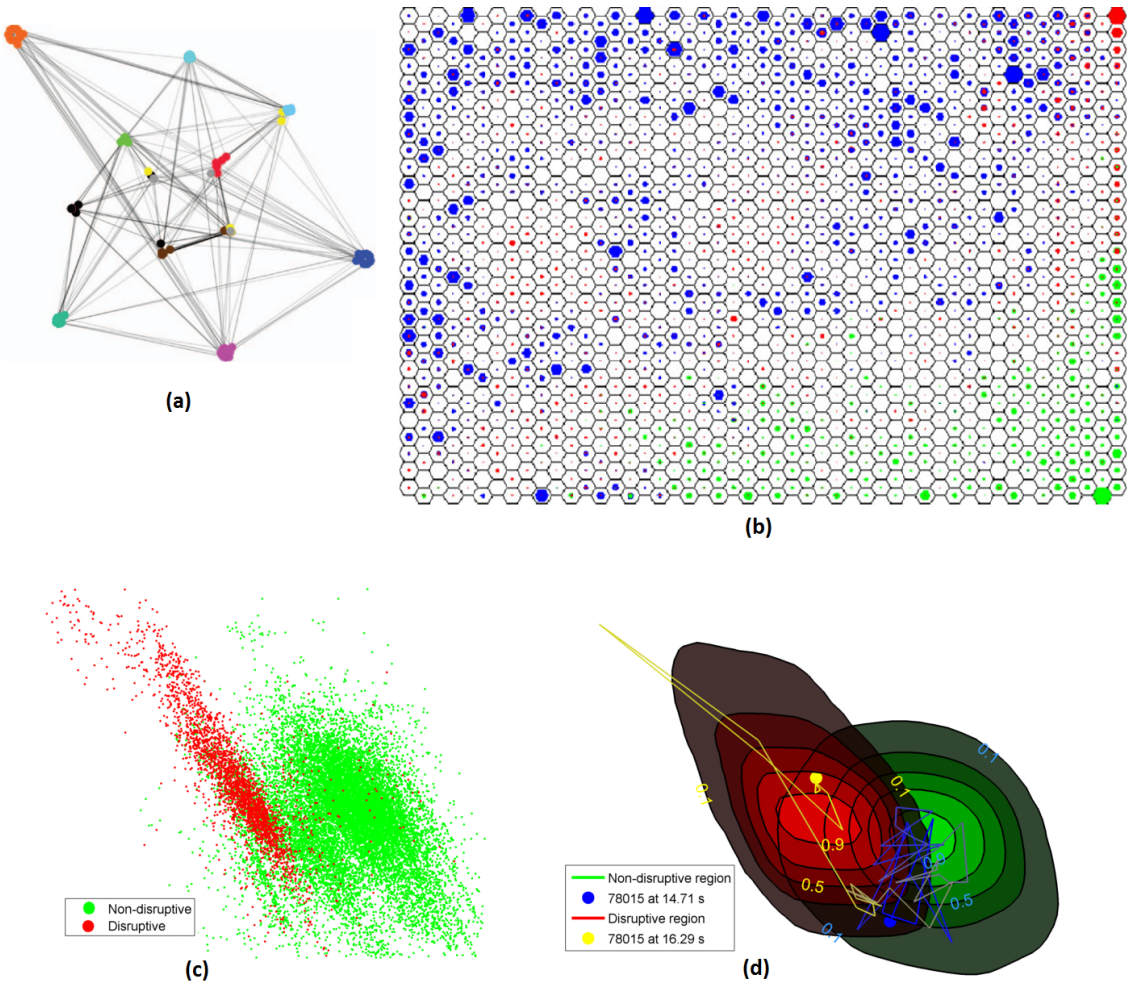


Figure 2.8: *Examples of data visualization using dimensionality reduction methods. (a) Visualization of a graph with the nodes indicting US college football teams and the edges representing the teams they played against. Figure adapted from [71]. (b) 2D map using self organising maps indicating safe (blue and red) and disruptive (green) clusters from 229 AUG discharges. For each cluster, the color density is proportional to the number of samples contained within the clusters. A transition region also appears between the safe and the disruptive regions. Figure reproduced from [72]. (c) and (d) are 2D projections from JET disruption data. (c). Disruptive and non-disruptive clusters of data points are mapped by landmark multidimensional scaling. (d). Density contours for the data points belonging to disruptive and non-disruptive clusters are drawn. The numbers give a qualitative measure of the degree of the disruptiveness of the region. A trajectory for pulse number 78015 is also traced.*

While dimensionality reduction facilitates data compression and storage, its use which is exploited in this thesis is *data visualization*.

Data visualization

Data visualization is an integral element of exploratory data analysis and an important first step in assessing the data before proceeding to specific modeling

and analysis. It is well known that a 2D embedding of high dimensional data for discovering meaningful information obscured by intrinsic data complexity is not entirely lossless. Hence a visualization method needs to select what kind of errors to make and the choice naturally should depend on the visualization goal. This has led to large array of data visualization techniques such as principle component analysis (PCA), generative topographic mapping (GTM) [73], self-organizing map (SOM) [74], multidimensional scaling (see section 2.3.1), to name a few. Some applications of data visualization are illustrated in Figure 2.8. Figure 2.8(a), is a 2D visualization of a graph using the latent variable model [71][75]. The nodes indicate the US college football teams and the edges depict the team they played against. The visual groups of teams match the 12 conferences arranged for yearly play. Figure 2.8(b) is a 2D map indicating the plasma states (safe or disruptive) based on 229 discharges from AUG from the experimental campaigns performed between July 2002 and April 2005. The mapping obtained via SOM, allows detection of the regions with a high risk of disruption [72]. In Figure 2.8 (c) and (d), the data acquired in disruptive shots at JET between campaign C21 and C27 has been used for obtaining visually informative plots for disruptions. Wavelet decomposition using 4 wavelet scales is carried out for 4 indicative signals, namely the I_p, ML, electron density n_e and P_{rad}. The distribution of wavelet coefficients is described by a zero-mean Laplace distribution which allows for a fast calculation of geodesic distances. Figure 2.8(c) shows the visualization obtained using landmark multidimensional scaling (see subsection 2.3.2) for disruptive and non-disruptive data points. Non-disruptive points are obtained from the indicator signals at 2 to 1 s before disruption, whereas the data points obtained from 210 to 30 ms preceding disruption constitute the disruptive points. Figure 2.8(d) provides a deeper, quantitative insight into the distribution of disruptive and non-disruptive plasma states by marking the density contours for the points in each cluster. The maximum density of points for each cluster is taken as the reference level 1, and the contours are defined with respect to the respective maximum level. Finally, a trajectory of discharge number 78015, a JET pulse that disrupted at 16.32 s due to the onset of a neoclassical tearing mode, is mapped. Figure 2.8(d) follows the trajectory from about 1.5 ss preceding the discharge.

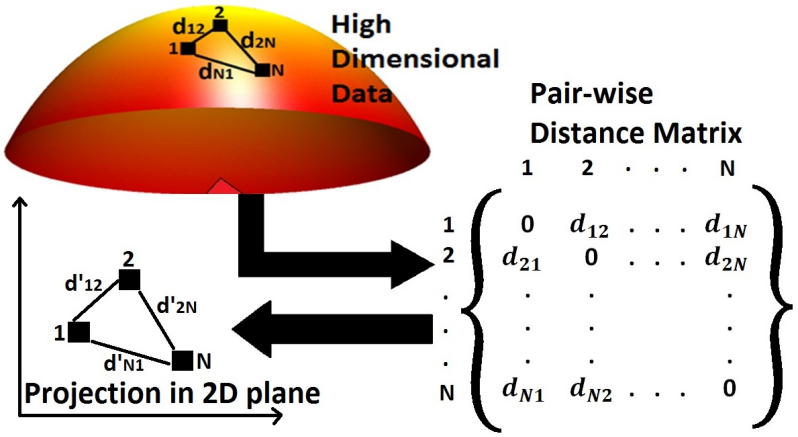


Figure 2.9: *Illustration of multi-dimensional scaling (MDS). High dimensional data points are projected in 2 dimensions, such that* $d'_{ij} \approx d_{ij}$ *where* d_{ij} *and* d'_{ij} *are distances between data points* i *and* j *in high dimensions and in 2D respectively.*

2.3.1 Multidimensional scaling

Classical multi-dimensional scaling (MDS) is a well-regarded dimensionality reduction technique used for providing a visual representation of a complex set of relationships [76][77][78]. It yields a projection in the 2D or 3D Euclidean plane of high-dimensional data, while ensuring minimal information loss during dimensionality reduction.

The working principle of MDS is illustrated in Figure 2.9. Let $X = 1, 2, ...N$ be a set of N (high) dimensional data points where the distance (or dissimilarity) between the ith and jth data point is d_{ij}. As shown in Figure 2.9, for a pair-wise distance matrix for high dimensional data points in X, MDS finds a lower dimensional mapping such that $d'_{ij} \approx d_{ij}$ for all data points, where, d'_{ij} is the distance between ith and jth data points in the lower dimensions. The following steps summarize the algorithm of classical MDS:

1. Matrix of squared distances (D) is set up: $D = [d^2]$.

2. Double centering is applied using the centering matrix J

$$B = -\frac{1}{2}JDJ \tag{2.14}$$

   where $J = I_N - N^{-1}\mathbb{O}$. I_N is the identity matrix of size N and $\mathbb{O}$ is an $N \times N$ matrix of all $1's$.

3. k largest positive eigenvalues $\lambda_1 ... \lambda_k$ of B and the corresponding k eigenvectors

$e_1 ... e_k$ are extracted.

4. A k-dimensional spatial configuration of the N data points is derived from the coordinate matrix $X = E_k \Lambda_k^{1/2}$, where E_k is the matrix of k eigenvectors and Λ_k is the diagonal matrix of k eigenvalues of B, respectively.

2.3.2 Landmark multidimensional scaling

Landmark multidimensional scaling (LMDS) is a computationally efficient approximation to the classical MDS [79]. Classical MDS has a complexity of (approximately) $O(kN^2)$ where N is the number of data points and k is the dimension of the embedding. This renders MDS practically unfeasible for very large datasets and necessitates the adoption of a better approach in place of an eigen decomposition of the full $N \times N$ matrix derived from the input distance matrix D.

LMDS as a computationally efficient variant of MDS is robust to noise and offers a correct mapping if the data really has a low-dimensional structure. The following steps summarize the algorithm of LMDS:

1. A set of n landmark points are randomly selected from the data set.

2. A $n \times n$ matrix D_n of distances between pairs of landmark points is computed for providing as input to step 3 below.

3. Classical MDS as outlined in section 2.3.1 is applied for finding a $k \times n$ matrix L representing an embedding of the n landmark points in k-dimensional space.

4. A distance-based triangulation (DBT) is then used for embedding the remaining data points in the k-dimensional space. DBT is a procedure through which a low-dimensional embedding is obtained by an affine linear transformation of the squared distances between the data and the landmark points.

It is noteworthy, that since in this doctoral work each data point is in fact a probability distribution, MDS and LMDS utilize the GDs between PDFs to create a powerful information visualization tool which is capable of yielding 2D maps for high-dimensional plasma data (see chapter 5)

2.4 Classification

2.4.1 Discriminant analysis

Discriminant analysis enables prediction (predictive discriminant analysis) of the class membership based on a linear or quadratic combination of input variables (such as plasma parameters). Secondly, it aids in the understanding of data, as a careful examination of the prediction model (descriptive discriminant analysis) that results from discriminant analysis can give insight into the relationship between class membership and the variables used for predicting class membership. It is a parametric method and assumes that the distribution within each class is multivariate normal.

Given two classes ($k = 1, 2$), each class k with n_k samples is denoted by a $(n_k \times p)$ data matrix where p is the number of plasma parameters. The class specific density of a sample x belonging to class $k = r$ is denoted as $f_r(x)$. Further, π_r denotes the prior probability of sample x belonging to class r, with $\Sigma_{r=1}^{k}\pi_r = 1$. The posterior probability of a sample x belonging to class $k = r$ is obtained by applying Bayes theorem:

$$P(r|x) = \frac{f_r(x)\pi_r}{\Sigma_{s=1}^{k} f_s(x)\pi_s}. \tag{2.15}$$

The denominator is consistent across all classes; hence it suffices to estimate class specific densities $f_r(x)$ for each of the classes. It follows that we classify x in class r if $f_r(x)\pi_r$ is maximal. Each of the class densities is modeled as a multivariate normal density:

$$f_r(x) = \frac{1}{(2\pi)^{p/2}\sqrt{|\Sigma_r|}} exp\bigg(-\frac{1}{2}(x-\mu_r)^t(\Sigma_r)^{-1}(x-\mu_r)\bigg). \tag{2.16}$$

The *Mahalanobis* distance of a sample x to class r is given as

$$d_r^2(x) = \sqrt{(x-\mu_r)^t(\Sigma_r)^{-1}(x-\mu_r)}. \tag{2.17}$$

CASE 1: Homoscedasticity

All classes are considered to be sharing a common covariance matrix. Hence, $\Sigma_r = \Sigma$

for all classes r. Taking the logarithm of $f_r(x)\pi_r$, we obtain for each class

$$\log(f_r(x)\pi_r) = \log f_r(x) + \log \pi_r. \tag{2.18}$$

Disregarding the constant term $p\ln(2\pi)$, we obtain

$$\log(f_r(x)\pi_r) = -\frac{1}{2}\log(|\Sigma|) - \frac{1}{2}(x-\mu_r)^t(\Sigma_r)^{-1}(x-\mu_r) + \log(\pi_r).$$

Removing the terms that are constant for all classes yields the class scores $l_r(x)$ given by

$$l_r(x) = x^t\Sigma^{-1}\mu_r - \frac{1}{2}(\mu_r)^t\Sigma^{-1}\mu_r + \log(\pi_r). \tag{2.19}$$

The score $l_r(x)$ is a linear function of x and this approach is called the *linear discriminant analysis* (LDA).. The decision boundary between two classes is the collection of points x for which $l_r(x) = l_s(x)$. In p dimensions the boundary between two classes is thus a hyperplane.

The class centers and the common covariance matrix for the classes are estimated from the training data. The standard estimates are:

$$\mu_r = \frac{1}{n_r}\sum_{k_i=r} x_i, \tag{2.20}$$

$$\Sigma = \frac{1}{n-k}\sum_{r=1}^{k}\sum_{k_i=r}(x_i-\mu_r)(x_i-\mu_r)^t. \tag{2.21}$$

Hence, the estimated centers μ_r for each class are the means of class samples and the covariance matrix estimate Σ is the pooled covariance matrix of the samples in all classes. Prior class probabilities also have to be estimated from the data and are given by:

$$\pi_r = \frac{n_r}{n}, \tag{2.22}$$

where n_r is the number of samples of class r and n is the total number of samples in the data.

CASE 2: Heteroscedasticity

Classes do not share a common covariance matrix. Similarly as before, taking the

logarithm $f_r(x)\pi_r$ and removing constants yields the class scores $q_r(x)$ given by

$$q_r(x) = -\frac{1}{2}\log(|\Sigma_r|) - \frac{1}{2}(x-\mu_r)^t(\Sigma_r)^{-1}(x-\mu_r) + \log(\pi_r). \tag{2.23}$$

The score $q_r(x)$ is a quadratic function of x. Hence, the decision boundary between any two classes $q_r(x) = q_s(x)$ is also quadratic. Therefore, this approach is called *quadratic discriminant analysis* (QDA). The covariance matrix for each class is estimated by the sample covariance matrix of the training samples in that class, that is

$$\Sigma_r = \frac{1}{n_r - 1}\sum_{k_i=r}(x_i - \mu_r)(x_i - \mu_r)^t. \tag{2.24}$$

[80][81]

2.4.2 k-nearest neighbor

The k-nearest neighbor classifier (kNN) [82] is a non-parametric, instance-based learning algorithm applied for incremental learning. The first formulation of a rule of the nearest-neighbor type was proposed in 1951 by Fix and Hodges [83], where they also gave a preliminary analysis of its properties. The underlying principle of the nearest-neighbor classification is that instances within a dataset will generally exist in close proximity to other instances that have similar properties. In order to classify the test/query sample, the nearest-neighbor algorithm finds its closest point in the d-dimensional training data $T = x_1, x_2, ..., x_n$ of n instances. The "closeness" or distance to the training data (neighbors) of an unclassified instance is determined by using a distance metric, such as the Euclidean distance. A survey of different distance metrics for kNN classification can be found in [84]. This framework can be extended to the k-nearest-neighbor case, in which k closest points in the training data are returned by the algorithm. The test sample is then classified by a majority vote of its neighbors and is assigned to the class most common amongst its k-nearest neighbors. For this reason, k is usually an odd (to avoid tied votes) positive number. The high degree of local sensitivity makes kNN highly susceptible to noise in the training data. Thus, the value of k may strongly influence the performance of the kNN algorithm. The optimal choice of k is a problem dependent issue, but techniques like cross-validation can be used to reveal the optimal value of k.

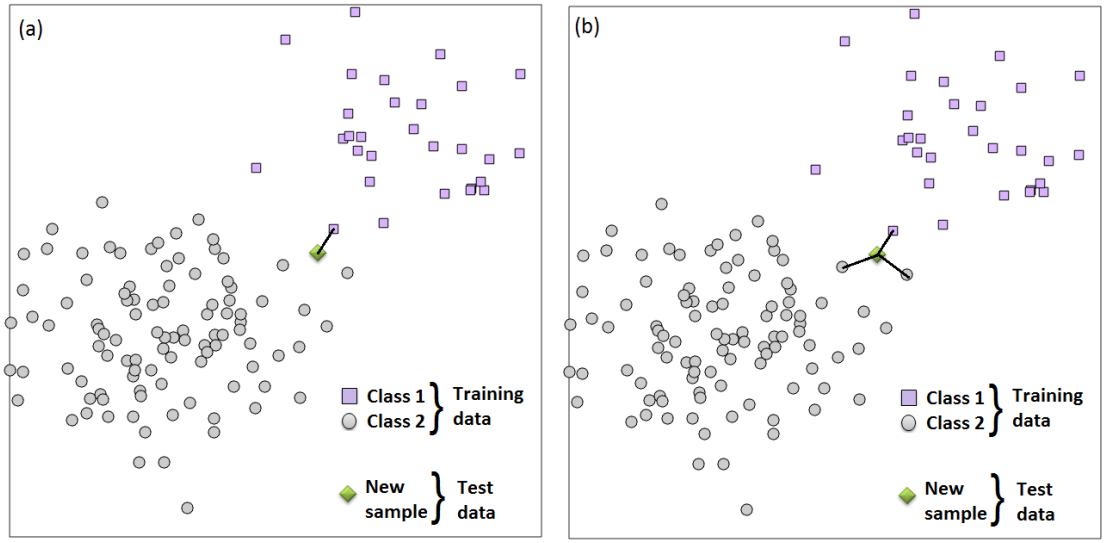


Figure 2.10: *Illustration of k-nearest neighbor in the Euclidean space. (a) Test sample is assigned class 1, using 1-nearest neighbor classification. (b) Test sample is assigned class 2 based on 3-nearest neighbor classification*

k-nearest neighbor ($k = 1, 3$) classification in the Euclidean space is illustrated in Figure 2.10(a) and (b) respectively. In Figure 2.10(a), the test sample is assigned to class 1 as its nearest neighbor belongs to class 1. In Figure 2.10 (b) the test sample is assigned to class 2 which is the majority class amongst its 3 nearest neighbors. On similar lines, k-nearest neighbor in spaces of probability distributions is illustrated in Figure 2.11. We know from differential geometry that a geodesic on a manifold is an analog of a straight line in the Euclidean space. In Figure 2.11, the test sample (a probability distribution) is assigned to class 2 as that is the dominant class amongst the 3 nearest neighbors of the test sample. The nearest neighbors are those instances from the training data which have the minimum GD with respect to the test sample.

kNN has a number of attractive properties. First, the asymptotic error rate of the 1-nearest neighbor classifier is never more than twice the Bayes rate as shown in the work by Cover and Hart [82]. Secondly, due to its non-parametric nature, it does not depend on knowing the form of the distribution from which the data has been drawn. Rather the inference can be made directly from the observed data i.e. there is no model building process. As a result of these characteristics, it has found application in numerous research fields [27][26].

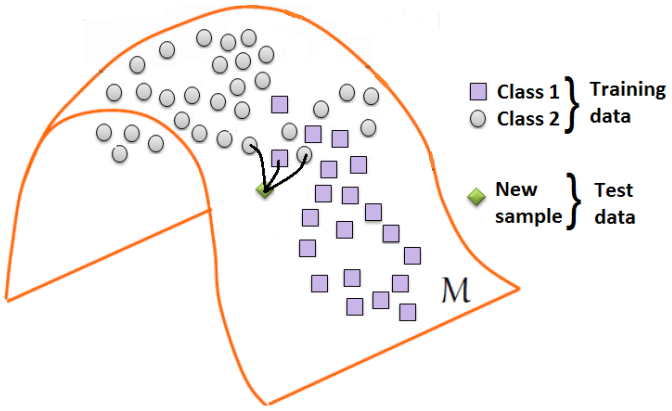


Figure 2.11: *Illustration of k-nearest neighbor on the manifold M. Test sample (probability distribution) is assigned class 2, using 3-nearest neighbor classification. The nearest neighbors are ascertained by computing GDs between the test sample and the samples in the training data.*

2.4.3 Conformal predictor

See, Conformal predictors (section 6.3)

2.4.4 Distance-to-centroid classifier

A *distance-to-centroid* (D2C) classifier, in principal, assigns a test sample to the class whose centroid is the closest (i.e. minimum distance) to the test sample. This is relatively straightforward in a Euclidean space where the centroid is simply the mean location of all the points, taken along each dimension separately. However, the manifestation of D2C in the manifold setting (Figure 2.12) is considerably more challenging.

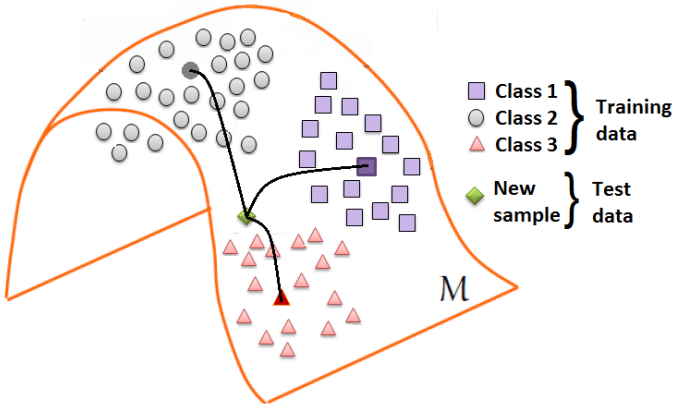


Figure 2.12: *Illustration of distance-to-centroid (D2C) classifier on the manifold M. Test sample is assigned class 3, as the centroid of class 3 is closest (minimum GD) to the test sample.*

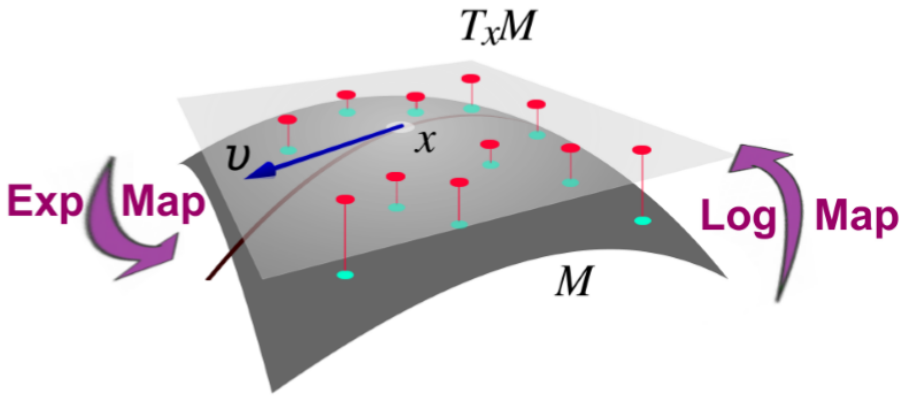


Figure 2.13: *Illustration of the tangent space T_xM to the manifold M at the point $x \in M$. Exponential map and logarithmic map for transforming to and from the manifold on the tangent space are also indicated.*

Fréchet mean

The Fréchet mean provides a generalization to the manifold setting of the centroid of a cluster of points in a Euclidean space. For a set of n points, x_i, $i = 1, ..., n$, the Fréchet mean μ_f can be obtained through the following minimization:

$$\mu_f = \underset{x}{\operatorname{argmin}} \sum_{i=1}^{n} GD(x, x_i) \tag{2.25}$$

This results in an optimization problem on the manifold which, assuming that a unique solution exists, can be solved by the gradient descent algorithm on the manifold [85]. The centroid is iteratively determined by projecting the points x_i on the tangent space at the initial Euclidean approximation of the centroid. The Euclidean mean in the tangent space is computed and the result is projected back to the manifold. This is illustrated in Figure 2.13. The transformation to and from the manifold onto the tangent space is achieved via an *exponential* map and a *logarithmic* map, respectively.

Exponential and logarithmic maps

In general, a d-dimensional Euclidean space $\mathbb{R}^d$ can be constructed at each point of a d-dimensional differentiable manifold M, tangent to M. The tangent space at a point $x \in M$ is denoted by T_xM. The exponential and logarithmic maps provide a local diffeomorphism between a differentiable manifold, M, and its tangent space centered at a point $x \in M$, T_xM. As illustrated in Figure 2.13, the transformation

from the tangent space to the manifold is called the exponential map and the inverse transformation is called the logarithmic map.

The steps for D2C classification are elaborated further in subsubsection 3.1.2.3.

2.4.5 Principal geodesic classifier

See principal geodesic classification (subsection 3.2.3).

Chapter 3

Texture discrimination

3.1 Multivariate texture discrimination based on geodesics to class centroids on a generalized Gaussian manifold*

A.Shabbir, G.Verdoolaege and G.Van Oost

Department of Applied Physics, Ghent University,
Sint-Pietersnieuwstraat 41, B-9000 Ghent, Belgium

Abstract. A texture discrimination scheme is proposed wherein probability distributions are deployed on a probabilistic manifold for modeling the wavelet statistics of images. We consider the Rao geodesic distance (GD) to the class centroid for texture discrimination in various classification experiments. We compare the performance of GD to class centroid with the Euclidean distance in a similar context, both in terms of accuracy and computational complexity. Also, we compare our proposed classification scheme with the k-nearest neighbor algorithm. Univariate and multivariate Gaussian and Laplace distributions, as well as generalized Gaussian distributions with variable shape parameter are each evaluated as a statistical model for the wavelet coefficients. The GD to the

*The work presented here in section 3.1, has been published in this form as:
A.Shabbir, G.Verdoolaege and G. Van Oost, "Multivariate texture discrimination based on geodesics to class centroids on a generalized Gaussian manifold," *Geometric Science of Information, ser. Lecture Notes in Computer Science*, vol. 8085, pp. 853-860, 2013.

centroid outperforms the Euclidean distance and yields superior discrimination compared to the k-nearest neighbor approach.

Keywords: Rao geodesic distance, texture discrimination, wavelet distributions.

3.1.1 Introduction

Web based browsing and digital image libraries have experienced an unprecedented growth in the last decade. A variety of texture classification and retrieval techniques have been developed for tackling the issue of automated discrimination of textured images and their subsequent retrieval, both online and offline. The major challenge in this application is the classification and extraction of the desired image with maximized accuracy and least computational load.

Texture classification is essentially a two-stage process: feature extraction and similarity measurement. Feature extraction entails the extraction of a minimalist set of features that accurately depict the image in question. The subsequent similarity measurement requires the determination of a distance function which gauges the similarity of images on the basis of their respective feature sets. These two stages essentially dictate the design and performance of the classification and then the subsequent retrieval system.

Various popular and widely acknowledged texture discrimination techniques deploy filtering or wavelet-like approaches for accomplishing texture classification or retrieval [86] [87]. Essentially, these techniques make use of the enhanced ease of modeling the information, when it is made available in a transformed domain. These approaches typically provide acceptable classification performances from large texture databases and are also endorsed by the physiological studies of the visual cortex which suggests that the wavelet decomposition is a natural way of image formation[88]. Moreover, representation by wavelet features enables the classification schemes to operate directly in the compressed domain as wavelets are the principal technology in image coding formats like JPEG. These significant advantages and reasonable success of various wavelet based texture classification schemes, motivates our choice of wavelet representation of textures for this work.

In this study, we have exploited a parametric probabilistic framework for yielding a precise and accurate descriptor of images and thus obviating the need of storing

or transmitting any redundant information. Numerous univariate models have been proposed for characterizing the wavelet subbands. Despite the ease of modeling and computation, these approaches do not completely exploit the rich texture information as they are inadequate for modeling the correlation between color bands. Multivariate distributions such as Generalized Gaussian [87] [48], Gaussian Scale Mixture [63] and alpha-stable distributions [6] have, also, lately been utilized with varying degrees of success, for modeling the spatial and/or color correlations of the wavelet coefficients.

In this work, we employ a singular probabilistic model for modeling both the texture and color information, contained in the images. Verdoolaege et al. [48] established that classification and hence retrieval performance improves if the information contained in the correlation between color bands is exploited. Extending on this notion, in our probabilistic framework we utilize a multivariate probability distribution for joint modeling of the spectral bands while assuming independence amongst the wavelet subbands corresponding to the same color. In this work, we initially make use of the univariate Gaussian, Laplacian and generalized Gaussian distributions as our statistical model, and we then subsequently deploy the multivariate Gaussian, Laplacian and Generalized Gaussian distributions for comprehensive modeling of the rich correlation between color bands prevalent in the textured images.

Once feature extraction has been accomplished through the imposition of a suitable statistical model on the wavelet detail coefficients of the textured image which is to be classified or retrieved, determination of a suitable distance or similarity measure remains the next pursuit. As numerous possibilities exist in terms of probabilistic models which can be utilized for modeling the wavelet detail statistics, there is also a wide variety in terms of distance measures that can be used for evaluating the distance between probability distributions. Euclidean distance, despite yielding acceptable performances in various textural retrieval contexts [59], is not a natural similarity measure between probability distributions [48]. Kullback-Leibler divergence (KLD) despite its popularity for evaluating similarities is in fact not a true distance measure. The Rao geodesic distance (GD), derived from the Fisher information, has been used in case of multivariate probability distributions and has outperformed KLD and Euclidean in many

contexts [48]. Furthermore, the GD is a natural similarity measure between probability distributions.

In this paper, we propose a new scheme for texture retrieval based on the calculation of the geodesic distance between the query image and the centroid of the texture classes. Furthermore, to provide an ease of reference, we compare the performance of our proposed scheme with the performance of the k-nearest neighbor classifier using the Euclidean distance. We also evaluate the outcomes of our proposed technique when it operates with Euclidean distance as the underlying distance measure. Initially we work with the grey-level textures generated from the luminance of the RGB color images and we then move on to full joint modeling of the wavelet coefficients corresponding to the three colour bands. We also examine the computational expense of our proposed classification technique. The rest of the paper is organized as follows: section 3.1.2 summarizes the statistical models, the Rao geodesic distance and our proposed texture classification scheme. The section 3.1.3 outlines the experimental setup and presents the attained classification results. Finally, section 3.1.4 concludes the paper.

3.1.2 Statistical modeling and similarity measures

3.1.2.1 (Multivariate) generalized Gaussian distribution

The multivariate generalized Gaussian distribution has been introduced in [81] for modeling the wavelet detail co-efficients. We present the univariate generalized Gaussian distribution, before proceeding to the multivariate case. The univariate generalized Gaussian distribution is given as:

$$f(x|\alpha,\beta) = \frac{\beta}{2\alpha\Gamma[1/\beta]}\, exp\Big[-(|x|/\alpha)^{\beta}\Big], \tag{3.1}$$

where Γ denotes the Gamma function and α and β are, respectively, the scale and shape parameter controlling the variance and the fall-off rate of the distribution. $\beta = 2$ yields the Gaussian distribution and $\beta = 1$, results in the Laplace distribution. We proceed to the multivariate generalized Gaussian distribution, defined in [48] as:

$$f(x|\Sigma,\beta) = \frac{\Gamma(\frac{m}{2})}{\pi^{\frac{m}{2}}\Gamma(\frac{m}{2\beta})2^{\frac{m}{2\beta}}}\, \frac{\beta}{|\Sigma|^{1/2}}\, exp\left\{-\frac{1}{2}\Big[X'\Sigma^{-1}X\Big]^{\beta}\right\}. \tag{3.2}$$

Here, m is the dimensionality of the probability space, and is equal to 3 in our case of colored images. The distribution reduces to a multivariate Gaussian case for $\beta = 1$ and to a multivariate Laplace case for $\beta = 0.5$. Γ is the dispersion matrix. Parameters for multivariate MGGD, Laplace and Gaussian were estimated using the method of moments, followed by maximum likelihood estimation [48].

3.1.2.2 Geodesic distance

The Rao geodesic distance in the context of information geometry provides an effective distance measure between probability distributions represented by points on a probabilistic manifold. Geodesic distances allow for length minimization on the probabilistic manifold and offer an edge in terms of data visualization that they enable on the manifold [89]. For fixed shape parameter β i.e. Laplace and Gaussian case, the geodesic distance between two MGGDs denoted by (β, Σ_1) and (β, Σ_2) is given in [48]:

$$GD(\beta, \Sigma_1 | \beta, \Sigma_2) = \left[\left(3b_h - \frac{1}{4} \right) \sum_i (r_2^i)^2 + 2 \left(b_h - \frac{1}{4} \right) \sum_{i<j} r_2^i r_2^j \right]^{1/2}, \tag{3.3}$$

with $r_2^i \equiv ln\lambda_2^i$ and λ_2^i, $i = 1, ..., m$, the m eigenvalues of $\Sigma_1^{-1}\Sigma_2$. In addition, b_h is defined by

$$b_h \equiv \frac{1}{4} \frac{m + 2\beta}{m + 2}.$$

With variable shape parameter there is no closed form for the GD and we used a linear approximation to the geodesic coordinate functions, to render the calculations computationally more feasible, see [55].

3.1.2.3 Distance-to-centroid classifier

We present a novel classification scheme for data points (i.e. textures in this application) expressed as probability distributions and laying as points on a probabilistic manifold. The scheme is outlined as:

– Training data is used for computing the centroid for each class of textured images. The geodesic centroid is calculated according to an iterative algorithm described in [90], based on a projection on the tangent space. To realize this, the (inverse) exponential map was calculated for each of the distribution

models used in this work.

– Distance is evaluated between the class centroids and each test data object (query image in this case), which is to be classified.

– Geodesic distance is used as the distance measure, due to its suitability as a natural distance measure between probability distributions.

– Test data object is assigned the class, whose centroid has the shortest geodesic distance to the object.

3.1.3 Classification experiments

3.1.3.1 Experimental setup

We carried out our experiments with grey-level and colored textures from a small dataset of 40 images from the Vistex database [91]. This is the same database that was used by Verdoolaege et al. [48] and Do et al. [7] for conducting wavelet-based texture retrieval. This enables a comparison with their results in the similar context. The database comprises of glimpses of different real-world natural scenes possessing sufficient homogeneity and having a 512 x 512 image size. Each image was divided into 16 128 x 128-sized non-overlapping subimages, yielding a database of 640 subimages. Furthermore, each subimage was expressed in the RGB color space. Grey-level images were generated from the original color images by calculating their luminance. Moreover, every color (or grey-level) component of each subimage was individually normalized to zero mean and unit variance resulting in the subimages from the same original image not generally lying in the same range, rendering the problem more challenging. Following this, a discrete wavelet transform was applied on every component with three levels using the Daubechies filters of length eight. The wavelet detail coefficients of every subband over the three color components (or the grey-level) were modeled by a (multivariate) Gaussian or Laplace distribution, or a generalized Gaussian distribution with variable shape parameter. The parameters of the probability models for all subbands constitute the feature set for a single subimage.

The classification experiment was implemented in two stages: training and testing. In the training stage, the class label of each image was assumed to be known,

which enabled the calculation of a centroid for each class. In the testing phase, the distance between the test image and the centroid of each class was calculated. The test image was then assigned the class, whose centroid had the smallest distance to the test image. Following that, we compared the assigned class label with the actual class label of the test image. We carried out the experiment repeatedly, using every subimage as a test image once. We finally calculated the average rate of successful classification as a performance measure. The experiments were conducted with the geodesic distance as a distance measure between the test image and the class centroid, and subsequently using the Euclidean distance. This way the GD could be compared as a similarity measure between probability distributions to the Euclidean distance.

In the last stage, the classification was also performed using the k-nearest neighbor classifier in conjunction with the Euclidean distance, to provide a reference for comparison of our proposed technique. When working with the k-nearest neighbor algorithm, we considered one of the 640 subimages to be a test image which is to be assigned to one of the 40 classes. The class labels of the other subimages were assumed to be known. Distance between the test image and each of the remaining images was determined and the test image was assigned to the class most common among the fifteen nearest neighbors of the test image. Choice of fifteen nearest neighbors is motivated by the hypothesis that the fifteen nearest neighbors of the test image should be the fifteen subimages originating from the same class to which the test image belonged. Following that, we compared the assigned class label with the actual class label of the test image. Again, we carried out the experiment repeatedly, using every subimage as a test image once. The correct classification rate was then assessed by calculating the ratio of the images that were correctly classified to the total number of images. We conducted the classification experiments initially on the grey-level equivalent of the 640 colour images and then we catered the corresponding full RGB colour images considering the complete correlation structure between the spectral bands. For each of these instances, as a statistical model for wavelet coefficients, we employed the multivariate Gaussian, Laplacian and generalized Gaussian, characterised by $\beta = 1$, $\beta = 1/2$ and variable β, respectively.

Classifier	Measure	Model	Gray Images	Colour Images
Distance-to-centroid	Geodesic	Gauss	0.04	0.3
		Laplace	0.042	0.33
		GGD	0.476	1.301
	Euclidean	Gauss	0.015	0.043
		Laplace	0.015	0.044
		GGD	0.034	0.094
k-Nearest Neighbour	Euclidean	Gauss	0.241	0.69
		Laplace	0.242	0.7
		GGD	0.55	1.5

Table 3.1: *Time, in ms, necessary for the classification of one textured image (nine wavelet subbands), using geodesic distance-to-centroid (D2C) classifier and the k-nearest neighbor (kNN) classifier, characterized by different models.*

3.1.3.2 Computational demands

Besides accuracy, computational load of a retrieval or classification technique is also a crucial yardstick of performance. Computational efficiency has a direct impact on the required resources and speed, and can be a limiting factor in various applications. We have measured the time taken by our proposed technique to classify a query image based on the distance to the class centroids and also the time taken for the k nearest neighbor algorithm to perform the same feat. The time taken for computation was measured on the same machine on which all calculations pertaining to this work were performed. The machine employed for this work was a Dell Precision T7600 equipped with an Intel Xenon(R) CPU at 2.4 GHz and 16 GB of RAM, running the 64-bit version of the Windows 7 operating system. The retrieval systems were implemented and run in MATLAB (version 8, R2012b, 64 bit) [92]. The durations are presented in Table 3.1. Classification conducted with the distance-to-centroid classifier, employing GD as the distance measure, takes considerably longer than the same classifier working with Euclidean measure. However, this is a direct consequence of accuracy-versus-speed trade-off, as the classification accuracy with GD clearly outperforms Euclidean with a large margin. The most noteworthy observation is the superior performance of our proposed distance-to-centroid classifier, compared to a k-nearest neighbor classifier, in terms of computational load. Distance-to-centroid proves to be computationally attractive, as it essentially reduces the comparisons required to correctly classify one query image, to the number of centroids, which are equal to the number of classes.

k-nearest neighbor, ideally, requires N number of comparisons to accomplish the same task, where N is the number of entries in the database. Laplace and Gaussian models consume less time for geodesic distances, in contrast to generalized Gaussian, which is a direct repercussion of the necessity of approximating the GD in the latter case.

3.1.3.3 Results and discussion

The results of our classification experiments on the Vistex database are presented in Table Table 3.2, followed by a discussion on significant observations. It can

Classifier	Measure	Model	Grey Images	Colour Images
Distance-to-Centroid	Geodesic	Gauss	83.59	97.17
		Laplace	84.38	97.81
		GGD	87.19	97.19
	Euclidean	Gauss	46.1	58.91
		Laplace	45.63	58.91
		GGD	42.03	50.00
k-nearest neighbor	-	Gauss	67.5	78.13
		Laplace	65.16	77.03
		GGD	55.47	47.97

Table 3.2: *Correct classification success rates (%), using different models for three wavelet scales, using D2C and kNN classifiers.*

be observed that when the correlation structure between the spectral bands is considered (i.e colour images), the classification accuracy is substantially enhanced in comparison to grey scale modeling. This however, happens at an escalated computational expense. The most significant result is the high classification accuracy, achieved with our proposed distance-to-centroid classifier based on GD, in contrast to the k-nearest neighbor classifier. Superior performance of distance-to-centroid classifier with GD, as opposed to Euclidean, further substantiates the worth of the GD as a well suited distance measure for probability distributions on a manifold. Finally, the GGD yields higher classification accuracy for grey-scale images with the GD-based distance-to-centroid classifier.

3.1.4 Conclusion and future work

In this paper we have proposed a new technique for classifying textures, when they are represented in the wavelet domain. We have shown the value of the Rao

geodesic distance as an efficient distance measure between probability distributions and hence, as an important aid to effective classification. We have also illustrated how texture classification can profit by exploiting the information residing in the rich spectral band correlation structure by joint modeling through multivariate distributions. Furthermore, we have applied various statistical models and hence we have showed their respective competences for accomplishing the task.

In the future, we envisage investigating the behavior of our developed technique and obtained conclusions on other data sets and applications. Furthermore, we plan to improve our classification technique by incorporating class variance, leading to the calculation of Mahalanobis distances on tangent spaces. Analyzing the effect of additive noise on the performance of our classifier is also aspired.

3.2 Multivariate texture discrimination using a principal geodesic classifier*

A.Shabbir[1,2] and G.Verdoolaege[1,3]

[1]Department of Applied Physics, Ghent University, B-9000 Ghent, Belgium
[2]Max Planck Institute for Plasma Physics, D-85748 Garching, Germany
[3]Laboratory for Plasma Physics – Royal Military Academy (LPP – ERM/KMS), B-1000 Brussels, Belgium

Abstract. A new texture discrimination method is presented for classification and retrieval of colored textures represented in the wavelet domain. The interband correlation structure is modeled by multivariate probability models which constitute a Riemannian manifold. The presented method considers the shape of the class on the manifold by determining the principal geodesic of each class. The method, which we call *principal geodesic classification*, then determines the shortest distance from a test texture to the principal geodesic of each class. We use the Rao geodesic distance (GD) for calculating distances on the manifold. We compare the performance of the proposed method with distance-to-centroid and k-nearest neighbor classifiers and of the GD with the Euclidean distance. The principal geodesic classifier coupled with the GD yields better results, indicating the usefulness of effectively and concisely quantifying the variability of the classes in the probabilistic feature space.

Index Terms- Texture classification, principal geodesic analysis, geodesic distance

3.2.1 Introduction

Several texture discrimination techniques have shown the wavelet representation to be a well suited domain for characterizing textures [59][48][80]. Hence, wavelet decomposition is often conducted for the generation of a set of features (*signature*) that accurately characterize the texture image. In many discrimination methods, each wavelet subband is modeled by a probability density function (PDF). The distribution parameters are estimated, composing the signature of the texture. The

*The work presented here in section 3.2 has been published in this form as:
A.Shabbir and G.Verdoolaege, "Multivariate texture discrimination using a principal geodesic classifier," in *Proc. IEEE International Conference on Image Processing*, pp. 3550-3554, 2015.

next step entails the use of an appropriate similarity measure for assessing the similarity of two textures based on their respective signatures.

The Euclidean distance (ED) and the Kullback-Leibler divergence (KLD) between probability distributions have yielded acceptable performances in various texture retrieval contexts [59][48]. However, the ED is not a natural similarity measure between probability distributions and the KLD is in fact not even a true distance measure. The Rao geodesic distance (GD) derived from the Fisher information has outperformed KLD and Euclidean in many contexts [48][80]. Therefore, in this work, the GD between multivariate probability distributions has been used, as it provides a natural similarity measure between PDFs.

Numerous univariate models, such as the generalized Gaussian [59] and Weibull [93], have been proposed for characterizing wavelet subbands. However, these models are inadequate for modeling the correlation between color bands and thus do not completely capture the rich texture information. In this work, we employ the multivariate Laplacian and Gaussian probability distributions for joint modeling of the spectral bands, while assuming independence amongst the wavelet subbands corresponding to the same color. Texture retrieval techniques frequently compute the distance between the unlabeled (query) texture image and the nearest texture in the training set [59][48][94] seldom taking into account the underlying shape and variability of the class. In this paper, we present a new scheme for texture discrimination based on the calculation of the minimum geodesic distance between the unlabeled texture and the principal geodesic (principal direction) for each class. The principal direction, also called the first 'principal component', of the class is the direction in which the class members exhibit most variance.

For data lying in Euclidean space, principal component analysis (PCA) [95] provides an efficient parameterization of class variability. It yields the principal components of the data corresponding to the eigenvectors of the data covariance matrix. However, in our proposed scheme the texture signatures are parameters of PDFs and are no longer elements of a Euclidean space but in fact constitute a Riemannian manifold. Hence, PCA, being a standard linear technique, cannot be applied to textures. Therefore we employ principal geodesic analysis (PGA) [96] to each class for determining the direction with the greatest variability on the manifold. PGA is a generalization of PCA for the manifold setting.

Further, we compare the performance of our proposed scheme with the performance of the GD-based k-nearest neighbor (kNN) [48] and distance-to-centroid classifiers [80] on the manifold. We also evaluate the outcome of the techniques when they operate with the ED as the underlying distance measure.

The rest of the paper is organized as follows: section 3.2.2 summarizes the statistical models and the Rao geodesic distance, section 3.2.3 presents our proposed principal geodesic classifier and section 3.2.4 outlines the experimental setup and presents the attained classification results. Finally, section 3.2.5 concludes the paper.

3.2.2 Multivariate texture modeling

3.2.2.1 The multivariate Laplace distribution

The multivariate Laplace distribution is a particular case of the multivariate generalized Gaussian distribution (MGGD) that has been introduced in [48] and [81] for modeling the wavelet detail coefficients for color images. The MGGD is defined in [48] as:

$$f(X|\Sigma,\beta) = \frac{\Gamma(\frac{m}{2})}{\pi^{\frac{m}{2}}\Gamma(\frac{m}{2\beta})2^{\frac{m}{2\beta}}}\frac{\beta}{|\Sigma|^{1/2}}\, exp\left[-\frac{1}{2}\Big[X(\Sigma)^{-1}X\Big]^{\beta}\right], \tag{3.4}$$

where $\Gamma(.)$ denotes the Gamma function and Σ is the dispersion matrix. β is the shape parameter and controls the fall-off rate of the distribution. Also, m is the dimensionality of the probability space, and is equal to 3 in our case of RGB colored images. The distribution reduces to a multivariate Gaussian case for $\beta = 1$ and to a multivariate Laplace case for $\beta = 0.5$. The parameters of the probability models are estimated via the method of moments followed by an optimization through maximum likelihood estimation [48].

3.2.2.2 Geodesic distance

The Rao geodesic distance (GD) between two multivariate Laplace or two multivariate Gaussian distributions denoted by (β, Σ_1) and (β, Σ_2) is given as:

$$GD = \left[\Big(3b_h - \frac{1}{4}\Big)\sum_i (r_2^i)^2 + 2\Big(b_h - \frac{1}{4}\Big)\sum_{i<j} r_2^i r_2^j\right]^{1/2}. \tag{3.5}$$

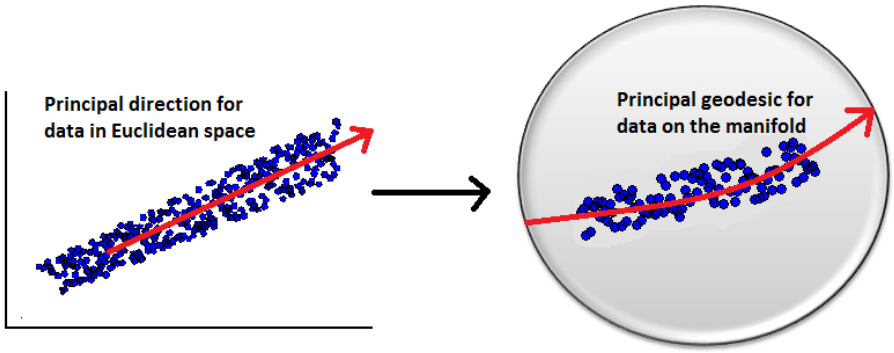


Figure 3.1: *The principal geodesic on a manifold is an analog of the principal component direction in the Euclidean space*

Here, $r_2^i \equiv ln\lambda_2^i$ and λ_2^i, $i = 1, ..., m$, are the m eigenvalues of $\Sigma_1^{-1}\Sigma_2$. Also, b_h is defined by,

$$b_h \equiv \frac{1}{4}\frac{m+2\beta}{m+2}.$$

3.2.3 Principal geodesic classification

A geodesic curve on a connected and complete manifold M is locally the shortest path between points. Essentially, a geodesic is a generalization of a straight line. Hence, a geodesic curve on the manifold is a natural analog of the first principal direction yielded by PCA. This is shown in Figure 3.1. PGA is outlined as follows:

- The class mean is computed for each class on the manifold. This entails the minimization of the sum of squared distance functions f for the class members $y_1, ...y_N \in M$

$$f(y) = \frac{1}{2N}\sum_{i=1}^{M} d(y, y_i)^2, \tag{3.6}$$

  This is achieved via a gradient descent algorithm first proposed by Pennec [97].

- The class members are now projected on the tangent space $T_\mu M$ of the manifold M at the class mean μ. The transformation to the tangent space is done through a *logarithmic map*:

$$log_\mu : y \in M \longrightarrow log_\mu(y) = \mu\vec{y}, \mu\vec{y} \in T_\mu M. \tag{3.7}$$

- PCA is conducted on the class members in the tangent space for obtaining the principal component directions (eigenvectors).

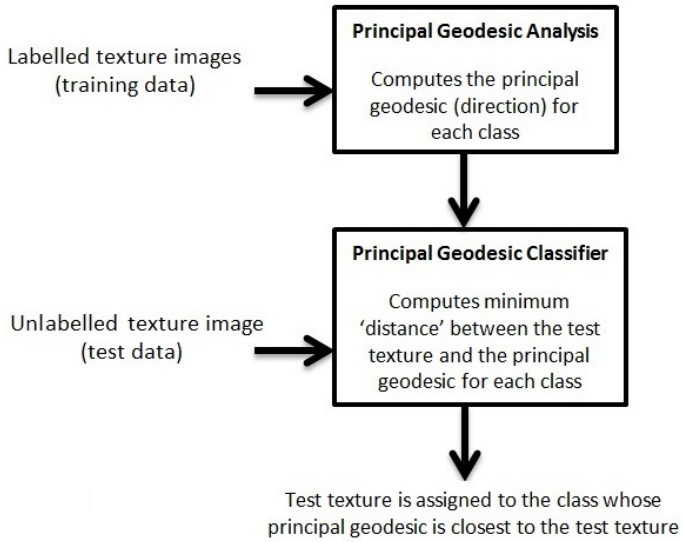


Figure 3.2: *Work-flow of principal geodesic classification*

- The eigenvector corresponding to the first principal component is projected onto the manifold using the *exponential map*:

$$exp_\mu : \mu\vec{y} \in T_\mu M \longrightarrow exp_\mu(\mu\vec{y}) = y, y \in M. \tag{3.8}$$

  This results in a point on the principal geodesic on the manifold.

The work-flow of principal geodesic classification (PGC) is given in Figure 3.2 In the training phase of the principal geodesic classifier, the principal geodesic is obtained for each texture class. In the testing phase, the distance of the test texture to the closest point on the principal geodesic is obtained via optimization (gradient descent) as shown in Figure 3.3. The test texture is assigned to the class whose principal geodesic is nearest to the test texture. Computationally, the advantage of this scheme is that only a few distances need to be evaluated in the gradient descent algorithm to find the distance to a specific class. This is opposed to e.g. kNN, which has to calculate distances to each sample in the database.

3.2.4 Classification experiments

3.2.4.1 Experimental setup

We carried out our experiments with 40 colored texture classes from the MIT Vision Texture (VisTex) database [91]. The database consists of glimpses of different natural scenes possessing sufficient homogeneity and having a 512 x 512 image

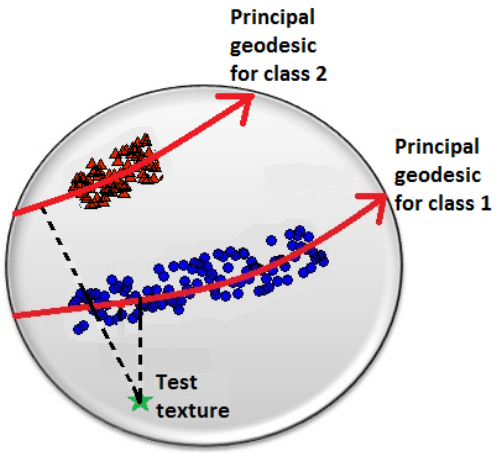


Figure 3.3: *Illustration of classification of a test texture by PGC. The distance of the test texture to the closest point on the principal geodesic is calculated for each class.*

size. From each of these texture images, 16 non-overlapping subimages of size 128 x 128 are created. This leads to a database of 640 subimages. Each subimage is expressed in the RGB color space. Further, every color component of each subimage is individually normalized to zero mean and unit variance resulting in the subimages from the same original image not generally lying in the same range. This renders the classification task even more challenging. Following this, a discrete wavelet transform with one level is applied individually on every component using Daubechies filter of length eight. The wavelet detail coefficients of every subband over the three color components are then modeled by a multivariate Gaussian or Laplacian distribution. These estimated parameters constitute the feature set for a single subimage. The dimensionality of the complete manifold is given by the number of independent entries in the dispersion matrices (6 for three-band color images), multiplied by the number of wavelet subbands.

In the training phase of the principal geodesic classifier, the principal geodesic for each class is computed assuming that the label for each texture image is known. 640 subimages are each used as a test texture once and their minimum distance to the principal geodesic of each class is calculated. Texture classification is also carried out using a distance-to-centroid classifier and kNN, to provide a reference for comparison with our proposed method. In the training phase of distance-to-centroid classifier, the centroid for each class is calculated. The test texture is assigned to the class whose centroid has the minimum distance to the test texture. Likewise, in kNN the test texture is assigned to the class most common amongst its fifteen nearest

neighbors. The choice of k =15 is driven by the hypothesis that the 15 nearest neighbors of the test texture should naturally be the 15 subimages originating from the same class to which the test texture belonged. Each subimage is treated as a test texture once, both in the distance-to-centroid classifier and kNN.

The correct classification success rate (%) for each classifier is then evaluated by calculating the ratio of textures that are correctly classified to the total number of textures.

The experiments are conducted with the GD as a distance measure and then also using the Euclidean distance (ED). This enables a comparison of the GD as a similarity measure between probability distributions to the ED.

3.2.4.2 Results

The results of the classification experiments on the VisTex database are presented in Table 3.3. The highest classification accuracy is achieved with our proposed principal geodesic classifier based on the GD, compared to distance-to-centroid and kNN. This indicates that accommodating the geometrical variability of the textures in the feature space can potentially lead to a performance improvement. PGA is essentially a dimensionality reduction procedure on the manifold, expressing each 6-dimensional texture image class by a single principal geodesic. This reduces the dimensionality of each wavelet subband to 1, yielding effective and concise image features. As mentioned before, PGC also offers a significant computational advantage over kNN. In addition, the superior performance of the classifiers with GD as a distance measure, compared to the Euclidean distance, further substantiates the superiority of the GD as a well-suited distance measure for probability distributions on a manifold. Finally, the Laplace distribution appears to be a better model than the Gaussian, though the differences in classification rates are marginal. On the other hand, it has been shown empirically in [48] that in retrieval applications the advantage of a Laplacian distribution can become more important. At this point it should be noted that, to the best of our knowledge, no analytic expression for the KLD between multivariate Laplace distributions has been found so far, as opposed to the GD.

Classifier	Measure	Model	SR (%)
Principal Geodesic	GD	Gauss	99.06
		Laplace	99.22
	ED	Gauss	71.25
		Laplace	75.00
Distance-to- centroid	GD	Gauss	95.94
		Laplace	95.78
	ED	Gauss	71.72
		Laplace	70.31
k-nearest neighbor	GD	Gauss	94.53
		Laplace	95.31
	ED	Gauss	69.06
		Laplace	69.53

Table 3.3: *Correct classification success rates (SR) (%), based on Laplace and Gaussian models for one wavelet scale, using principal geodesic, distance-to-centroid and k-nearest neighbor classifiers.*

3.2.5 Conclusions

In this work, we have presented a new texture discrimination method and demonstrated its classification performance on a database of 640 textured images. The presented principal geodesic classifier performs better than distance-to-centroid and k-nearest neighbor classifiers, making use of a highly optimized set of features on a probabilistic manifold. Further, we have shown the superior classification performance of the GD versus Euclidean distance in all our experiments.

Investigating the performance of our proposed classifier on other data sets and applications will be a subject of future work.

Chapter 4

Edge-localized modes in tokamak plasmas

In this chapter, the general introduction on the high confinement regime (H-mode) is completed. This is followed by a description of the plasma edge instability that is ubiquitous in H-mode regimes and is the prime focus of analysis in this thesis: the edge localized modes (ELMs). ELM phenomenology, physics and their control schemes are then briefly reviewed.

4.1 The H-mode and edge-localized modes

As already introduced in subsection 1.3.2, auxiliary heated plasmas undergo a transition from the L-mode to the H-mode which occurs as a bifurcation exhibited as a sudden increase in particle and stored energy confinement. In Figure 4.1, a standard H-mode discharge from AUG and a carbon-wall discharge from JET are shown, with the L-H transition being indicated with a dashed line. It can be observed that after the L-H transition, the electron temperature (T_e) (keV), density (n_e)($10^{-19}m^{-2}$) and stored energy (W_{MHD}) (MJ) increase rapidly. However, the increase in stored energy in the H-mode is not only due to the increase in heating power, but also due to a longer energy confinement time (τ_E) (s). τ_E is defined as

$$\tau_E = \frac{W}{P_{input}}, \tag{4.1}$$

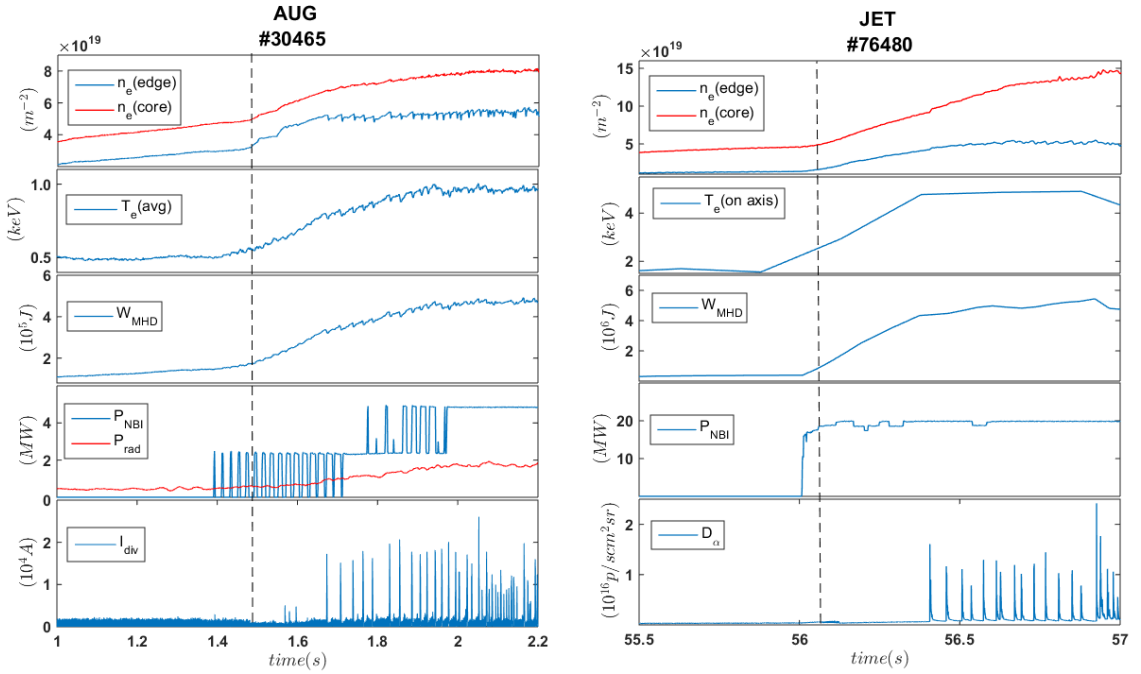


Figure 4.1: *L-H transition in a standard H-mode AUG discharge #30465 and JET carbon wall discharge #76480. With increasing input power (P_{NBI}) the electron temperature T_e, density n_e and W_{MHD} increase as well. The rate of increase becomes faster after the L-H transition which is indicated by the dashed line.*

and

$$W = \frac{3}{2}\int p\, dV, \tag{4.2}$$

where W (MJ) is the plasma stored energy, and P_{input} $(MW)$is the net input power (ohmic and auxiliary) that is delivered to the plasma. As shown in Figure 1.4, the improved confinement of the H-mode is due to the formation of an edge transport barrier (ETB) just inside the separatix in diverted tokamaks. The transport barrier is an outcome of a sheared perpendicular rotation profile, driven by a radial electric field, $E_r \times B$, in the plasma edge, which suppresses the long wavelength turbulence of the L-mode. This is illustrated in Figure 4.2. The transport barrier created by the so called E_r well gives rise to a steep gradient in the edge pressure profile. However, the steep plasma pressure gradient and the associated increased current density at the edge pedestal could exceed a critical threshold value and drive magnetohydrodynamic (MHD) instabilities called *edge localized modes* (ELMs).

ELMs are short, repetitive relaxations of the pedestal profiles which lead to particle and energy loss [98][20][19][99]. The edge pedestal collapses towards a shallower pressure gradient during the first stage of the ELM cycle namely the ELM *crash*. Figure 4.3 shows the evolution of several physical quantities within the ELM

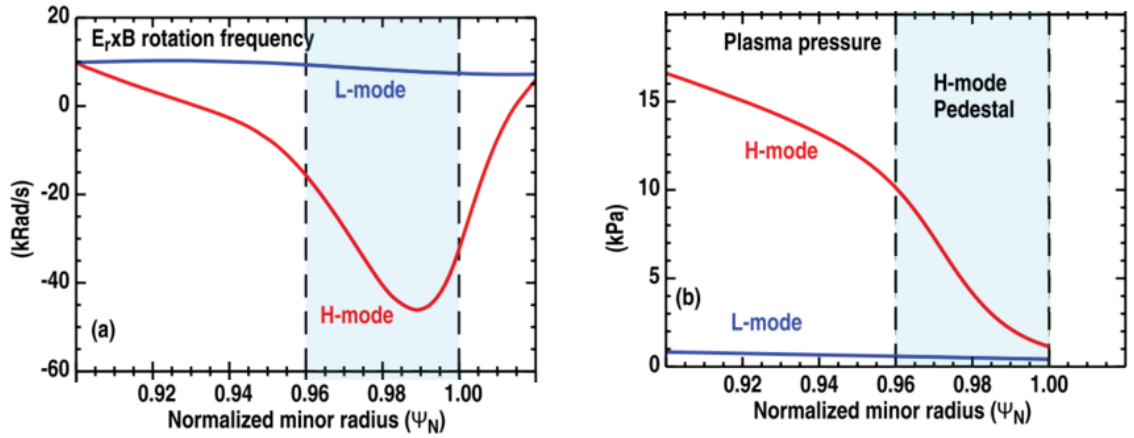


Figure 4.2: *(a) The edge perpendicular rotation profile ($E_r \times B$) in the H-mode and the L-mode. (b) The edge pressure profile in the H-mode and the L-mode. Figures adapted from [98]*

cycle for a standard H-mode discharge from AUG and a CW discharge from JET. Historically, the increase in the radiation in the D_α signal in the divertor, as shown in Figure 4.3(b), has been used as the ELM marker. The increase in D_α indicates an increase in edge recycling and can be used to give a measurement of the inward particle flux. However, since the installation of the full W divertor at AUG and

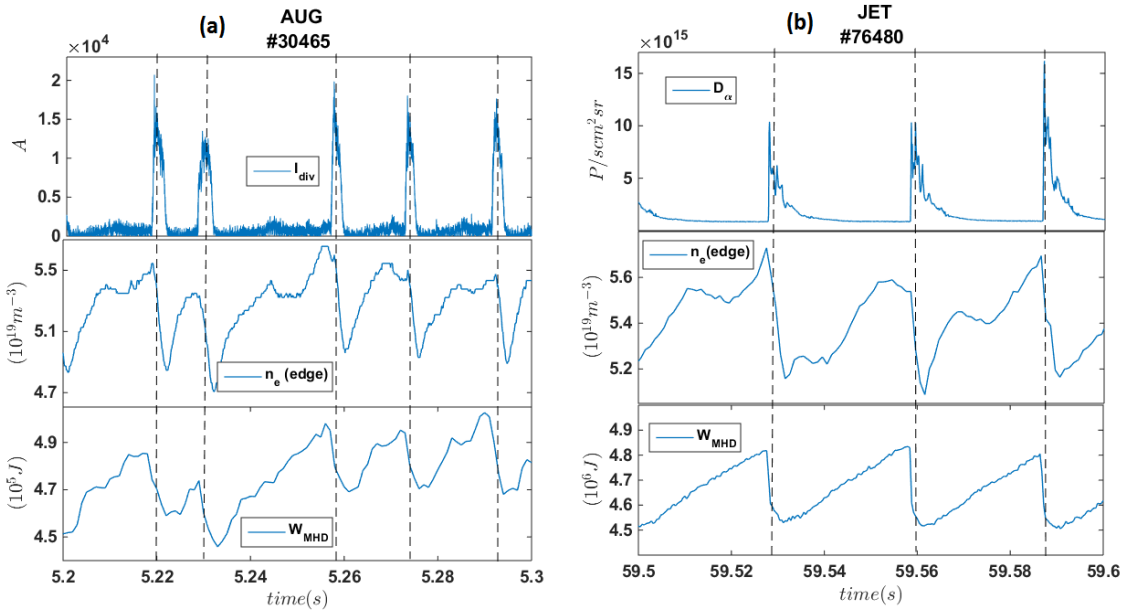


Figure 4.3: *Evolution of several physical quantities within the ELM cycle. (a). A standard H-mode AUG discharge #30465. (b). JET CW discharge #76480. In (a) current to the outer divertor I_{div}, electron density $n_e(edge)$ integrated along the edge interferometer line of sight inside the confined plasma and thermal plasma energy W_{MHD} obtained from equilibrium reconstruction are shown. In (b) intensity of the D_α radiation alongside $n_e(edge)$ and W_{MHD} are shown. Vertical dashed lines in both (a) and (b) indicate the ELM crashes.*

the ITER-like wall in JET, the quality of this signal has degraded. A high quality alternative at AUG, as shown in Figure 4.3(a), is the scrape-off layer (SOL) current measured via shunt resistances measured through the metallic divertor. Likewise at JET, with the ILW, the $Be\ II$ (527nm) fast emission signal in the divertor gives a better quality ELM signature.

ELM crashes in Figure 4.3 are identified by vertical dashed lines. After these crashes all quantities recover on a slower time scale towards their pre-ELM values. This constitutes the second part of the ELM cycle referred to as the *build-up* phase or the *recovery* phase. ELMs can reduce the pedestal energy confinement by ∼10-20% as well as lead to large transient heat and particle loads on the PFCs. ELM energy loss (W_{ELM}) is typically normalized by the pedestal energy (W_{ped}) when comparing relative ELM sizes across devices of different size. Under comparable conditions the relative ELM size, W_{ELM}/W_{ped}, is roughly constant for a range of device sizes [100].

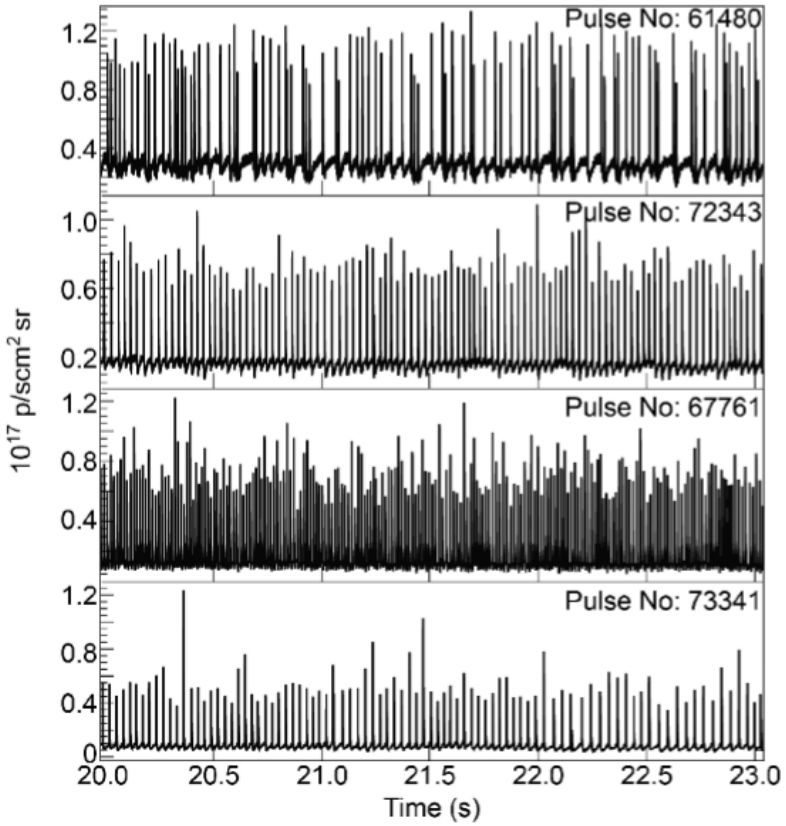


Figure 4.4: *D_α time trace showing type I ELMs in JET pulse number 61480 ($f_{ELM} = 33Hz$), 72343 ($f_{ELM} = 33Hz$), 67761 ($f_{ELM} = 67Hz$) and 73341($f_{ELM} = 29Hz$). Figure adapted from [102].*

4.2 ELM types

The physical mechanisms for the different ELM phenomena are complex and ELMs are classified from an empirical and phenomenological perspective [99][19][20]. A first classification was presented by E.Doyle *et al.* for the DIII-D tokamak [101] wherein three types of ELMs were identified and numbered according to the chronological sequence in which they were found. In general, there are three main criteria that have been used for ELM classification:

- Dependence of ELM repetition frequency on the heating power (the energy flux through the separatix)
- Occurrence of magnetic precursors, and
- MHD stability analysis, though Zohm [20] shows that the ideal ballooning stability is not a good criterion for separating the different ELM types.

4.2.1 Type I ELMs

Type I ELMs are characterized by an ELM repetition frequency (f_{ELM}) that increases with the energy flux through the separatix:

$$\frac{df_{ELM}}{dP_{sep}} > 0, \tag{4.3}$$

As shown in Figure 4.4, they typically appear as large, isolated bursts in the $D_\alpha/BeII$ emissions and are therefore also referred to as 'large' or 'giant' ELMs. These ELMs occur in good confinement regimes but result in an expulsion of a large amount of energy. An international multi-device database assembled by the ITPA Divertor/SOL Topical Group indicates that relative W_{ELM} is strongly correlated with pedestal collisionality (ν^*_{ped}) [103][104] and large ELMs can lose up to 20% of the W_{ped} at low values of ν^*_{ped} that are foreseen for ITER. Here ν^*_{ped} is defined as

$$\nu^*_{ped} = R\, q_{95}\, \varepsilon^{-3/2} (\lambda_{e,e})^{-1} \tag{4.4}$$

where, $\lambda_{e,e}$ is the mean electron-electron Coulomb collision mean free path length.

For the database in [103], the variation of relative W_{ELM} with ν^*_{ped} is shown in

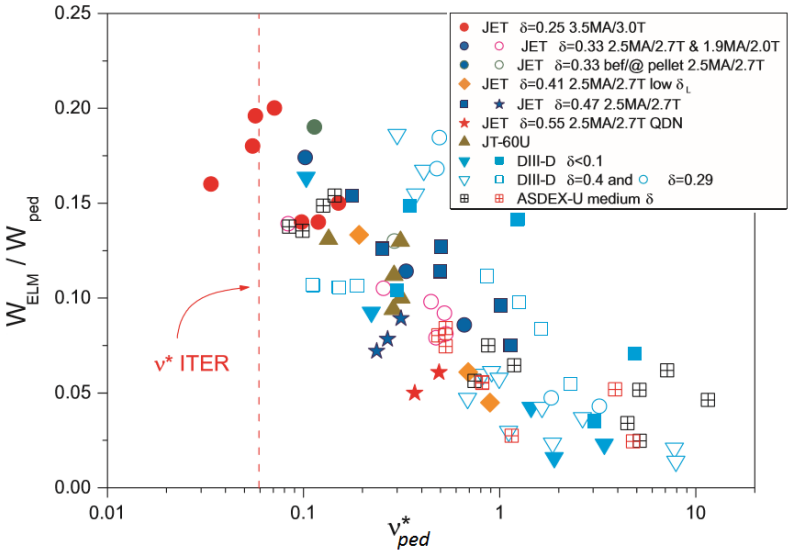


Figure 4.5: *Normalized ELM energy loss W_{ELM}/W_{ped} versus pedestal plasma collisionality (ν^*_{ped}) for a large range of type I ELMy H-mode plasmas in AUG, DIII-D, JT-60U and JET. Figure reproduced from [103].*

Figure 4.5. It can be noted from Figure 4.5 that while ν^*_{ped} is certainly a sorting parameter for ELM size in type I ELMy H-mode plasmas, there is, nevertheless, a significant scatter in the ELM size at a fixed ν^*_{ped}. On the one hand, this can partially be attributed to variations in the other parameters which influence W_{ELM} such as the plasma shape, heating levels, pedestal width etc. On the other hand, it is noteworthy that ELM size can vary significantly for ELMs within a single discharge. For type I ELMs at JET, Loarte *et al.* [107] measured a 15% standard deviation

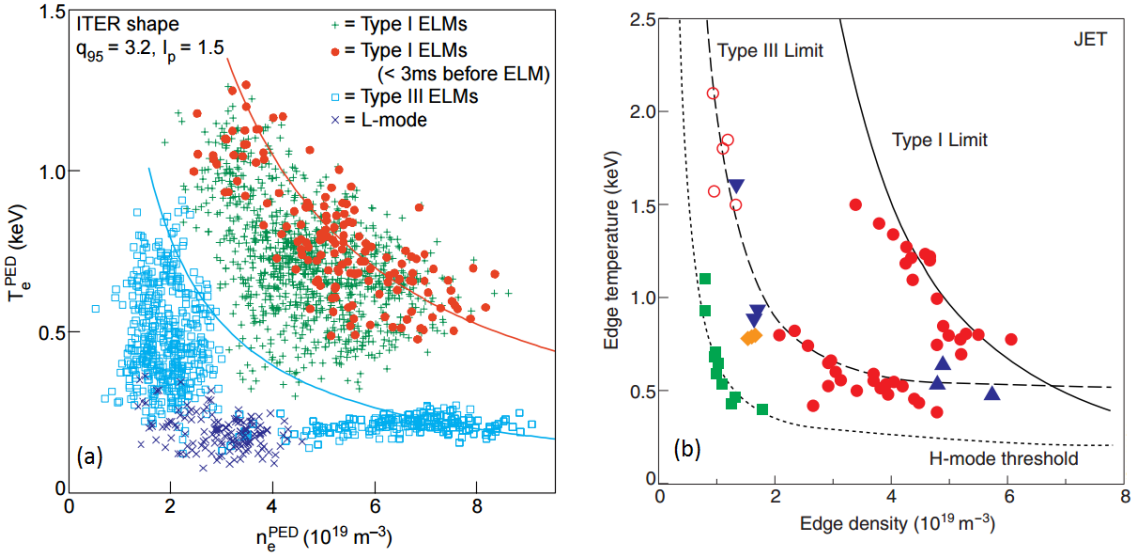


Figure 4.6: *Occurrence of ELM types in edge n_e–T_e space for (a). DIII-D (b). JET. Figure reproduced from (a). [105] (b). [106].*

in W_{ELM} for individual ELMs under stationary conditions. Webster *et al.* [108] in their analysis of ∼10,000 statistically equivalent ELMs from JET (B_t = 2T, I_p = 2MA, P_{NBI} = 12 MW, $\Gamma_{D_2} = 1.4 \times 10^{22} s^{-1}$, δ=0.2) note ∼50% standard deviation around the average value of W_{ELM}.

The onset of type I ELMs generally take the form of a critical edge pressure corresponding to a hyperbolic curve in the n–T space. Examples from DIII-D [105] and JET[106] are illustrated in Figure 4.6. It can be noted in Figure 4.6 that ELM types are well separated in the n–T space.

4.2.2 Type II ELMs

Type-II ELMs [101][20] have been observed only in strongly shaped plasmas (high elongation and triangularity) and at high collisionality [109] [110]. A simple criterion such as the dependence of f_{ELM} on input power has not been found for type II ELMs. In contrast to type I ELMs, there is an enhanced magnetic turbulence in the inter-ELM phase and the large periodic spikes on the D_α are substituted by small and more irregular oscillations. With the magnitude of ELM bursts lower than type I ELMs and the frequency higher, type II ELMs exhibit potential for a steady-state tokamak operation with good confinement accompanied with efficient impurity exhaust and tolerable loads on the PFCs. However, they appear in a narrow operational window and it is still unclear whether they will be possible to achieve in a burning fusion plasma.

4.2.3 Type III ELMs

Type III ELMs are characterized by an ELM repetition frequency (f_{ELM}) that decreases with the energy flux through the separatix:

$$\frac{df_{ELM}}{dP_{sep}} < 0. \tag{4.5}$$

On the D_α time trace, type III ELMs appear as small and frequent bursts. They typically expel 1-5% of plasma energy (W_{plasma}) which is significantly lower than the W_{ELM} associated with type I ELMs. However, the overall rather high energy transport leads to a stronger degradation of the energy confinement of the plasma compared to other ELM types. They are preceded by a coherent magnetic precursor

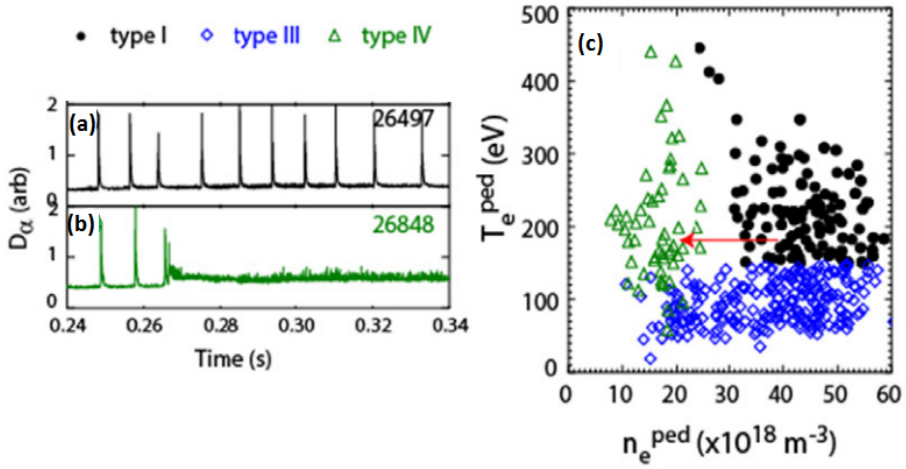


Figure 4.7: *Type IV ELMs at MAST obtained after mitigation of Type I ELMs using RMPs. (a). D_α time trace for target type I ELMs. (b). D_α time trace for type IV ELMs obtained after application of RMPs. (c). $n_{e_{ped}} - T_{e_{ped}}$ space as a function of ELM types, constructed from profiles obtained in the last 10% of the ELM cycle. Figures adapted from [112].*

oscillation of toroidal mode number $n \approx 5$–10 and poloidal mode number $m \approx 10-15$ [20][111]. In the n-T space, as shown in Figure 4.6, they are seen to occur below a much lower pedestal pressure and can be divided into two clusters: one at low $T_{e_{ped}}$ and high $n_{e_{ped}}$ and the other at high $T_{e_{ped}}$ and low $n_{e_{ped}}$.

4.2.4 Other ELM types

In addition to aforementioned three conventional ELM types, there are still other different ELM types, such as:

- **Type IV ELMs**: The low $n_{e_{ped}}$ and high $T_{e_{ped}}$ branch of type III ELMs (Figure 4.6(a)) is often termed type IV ELMs. Figure 4.7 (c) presents the D_α time trace for type IV ELMs ($f_{ELM} \sim 2000\ Hz$) at MAST that are obtained after the application of resonant magnetic perturbations (RMPs) to type I ELMs such as those shown in Figure 4.7(b) ($f_{ELM} \sim 2000\ Hz$). The n-T space diagram measured by a Thomson scattering system, shows the change of pedestal characteristics from a region typically associated with type I ELMs to one associated with naturally occurring type IV ELMs [112].

- **Type V ELMs**: Type V ELMs have been observed in high performance regimes at NSTX (National Spherical Torus Experiment)[113][114]. They are characterized by a short-lived $n = 1$ magnetic precursor oscillation rotating

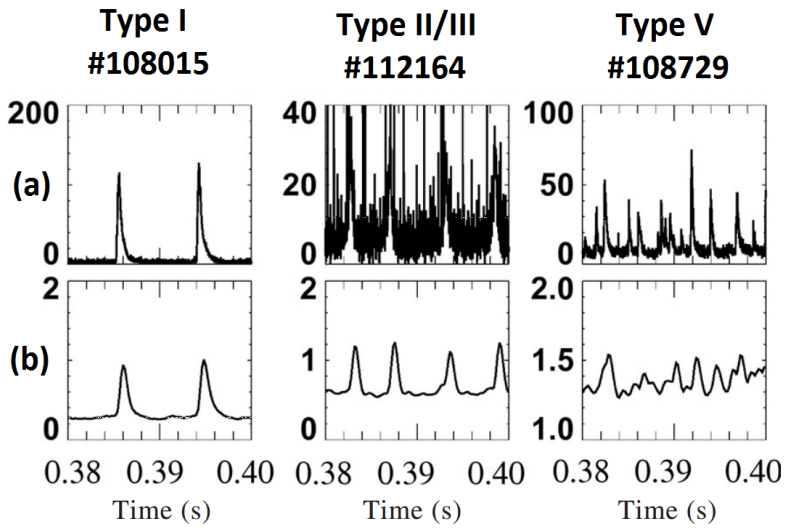


Figure 4.8: *Characteristics of different ELM types in NSTX. (a). ultra soft X-ray (b). lower divertor D_α emission. Figure adapted from [113].*

counter to the plasma current and are observed over a wide range of heating power in lower single-nulls. The signature of type V ELMs at NSTX is illustrated in Figure 4.8. To facilitate an ease of comparison, type I ELMs and a mixed type II/III ELMs signature at NSTX is also shown in Figure 4.8.

- **Grassy ELMs**: The grassy ELM regime has been found in JT-60U at lower collisionality in high poloidal beta (β_p) plasmas with a high triangularity (δ) [115][116]. The regime can be extended to lower $q95$ ($q95 < 4$), by increasing δ up to 0.6 [117]. Oyama *et al.*, as shown in Figure 4.9, describe grassy ELMs as having high frequency periodic spikes with small amplitude in the divertor D_α signal. They note that grassy ELMs like type I ELMs and in contrast to type III ELMs, obey a linear relation between f_{ELM} and P_{sep}. On JET and AUG, grassy-like ELMs have been observed in H-mode plasmas with $\beta_p > 1.7$, $q_{95} \sim 7$ and $\delta > 0.4$ [110][118] and have some synonymy with type II ELMs.

- **Compound ELMs**: At the transition from type I to type III ELMs as well as in steady state type I ELMy H-mode plasmas, a "compound" structure is often observed in the D_α time trace right after a type I ELM crash. This ELMy phenomenon at JET is referred to as compound ELMs. As can be noted from Figure 4.10, the large type I ELM spike in D_α is followed by a short interval of more frequent ELMs with a lower amplitude. Compound ELMs are most

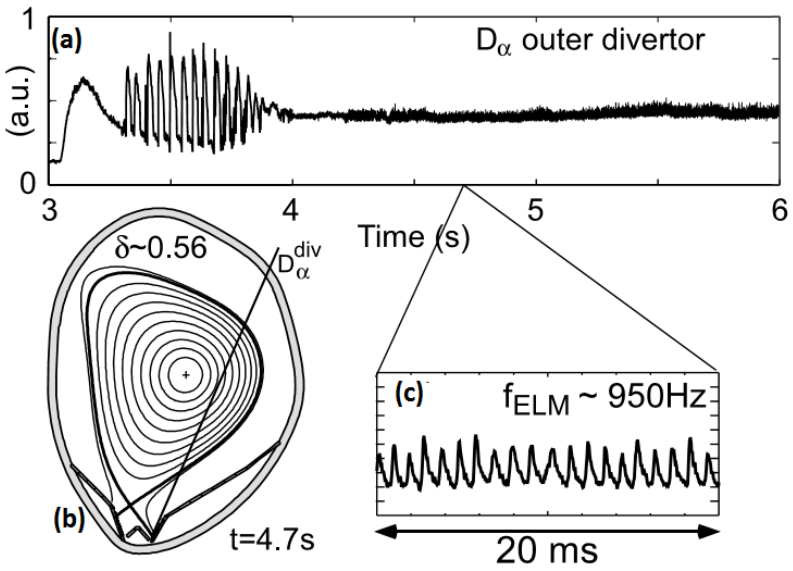


Figure 4.9: *A typical grassy ELM discharge at JT-60U with I_p= 1.0 MA and B_t = 3.9 T. (a). D_α at outer divertor. (b). Plasma configuration along with the line of sight for the D_α signal at outer divertor. (c). Magnified view of the D_α signal shown in (a) at t=4.7s. Figures adapted from [116].*

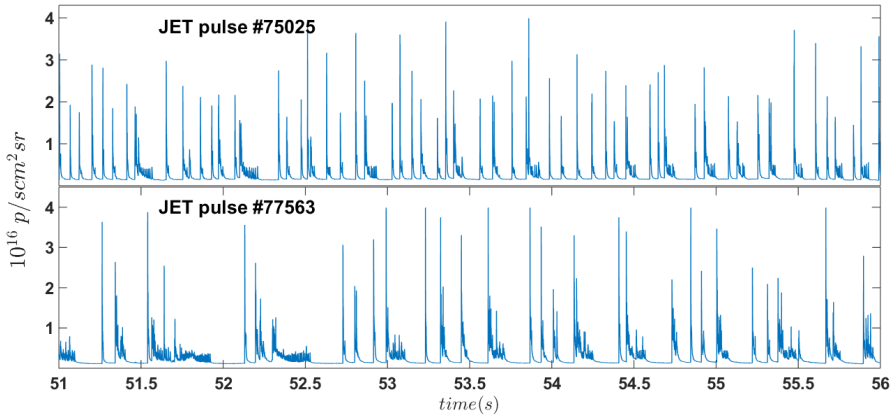


Figure 4.10: *D_α time traces for compound ELMs at JET.*

likely short periods of L-mode phase or are type III ELMs that follow a type I ELM, even though the power dependence of f_{ELM} is difficult to verify as compound ELMs occur in a narrow power range and under non-stationary conditions.

- **Dithering cycles**: Zohm [20] and Connor [111] enlist dithering cycles as an edge localized phenomenon. For heating power close to the H-mode threshold (P_{LH}), dithering cycles are repetitive L-H-L transitions. They show no magnetic precursor oscillation and exhibit a repetition frequency that decreases slightly with increasing P_{sep}.

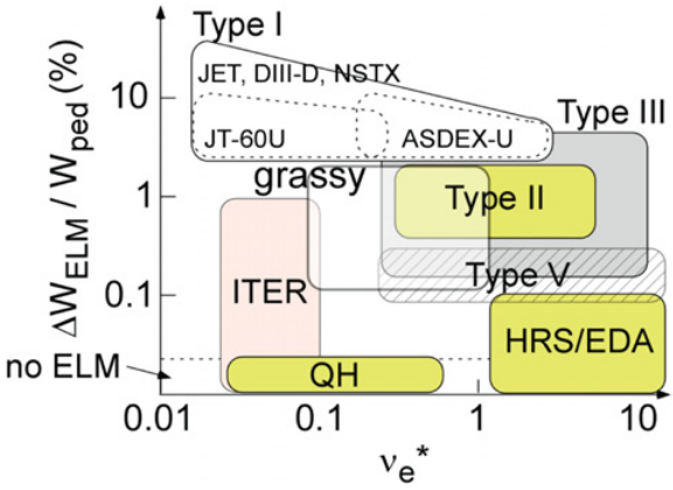


Figure 4.11: *Normalized ELM energy loss (W_{ELM}/W_{ped}) versus pedestal plasma collisionality (ν^*_{ped}) for several ELM types and small/no ELM regimes. Figure reproduced from [119]*

Additionally, several small/no ELM regimes such as EDA (enhanced D_α H-mode) [120][121], HRS (high recycling steady)[122], QH-mode (quiescent H-mode)[123] and I-mode [124][125] with good confinement properties have been obtained in Alcator C-Mod, AUG, DIII-D, JET, JFT-2 M, JT-60U and NSTX. An overview of the operational space achieved for several regular, small and ELM-less regimes in terms of the normalized ELM energy loss and edge collisionality is presented in Figure 4.11. ITER $Q \sim 10$ scenario, however, is based on controlled type-I ELMy H-mode regime and it remains unclear whether these small/ELM less regimes can be accessed on ITER. Characteristics and extrapolation of small/ELM less regimes, nonetheless, remains an active area of ongoing research.

4.3 Theory of ELMs

ELMs were identified as an ideal MHD instability soon after their discovery in the early 1980's [126]. Since then, considerable theoretical work has been performed, both analytically and through modeling calculations, for improving their theoretical understanding. Three types of ideal MHD instabilities can be expected at the ETB [127]:

- **Kink/Peeling modes:** The peeling mode is a specific kind of an external kink mode. It is destabilized by the finite edge current density and is dependent on the location of the closest rational surface to the plasma edge in the vacuum.

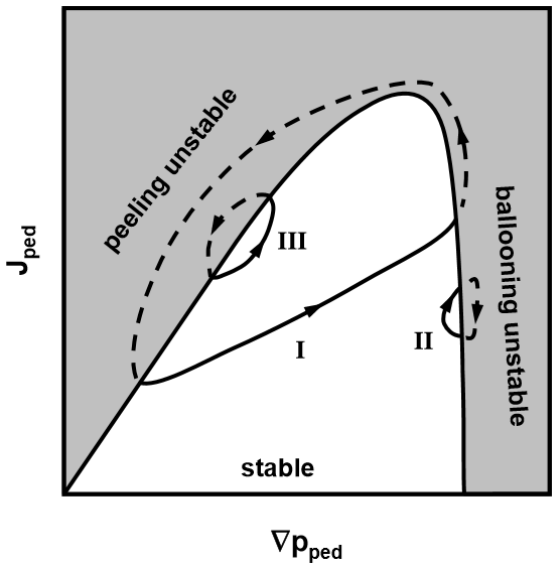


Figure 4.12: *ELM stability diagram, based on ideal MHD instabilities. For type-I ELMs, the edge instability is both pressure and edge current density driven and close to the corner of the peeling-ballooning stability. Type II ELM instability is pressure driven and close to the ballooning limit, while type III ELM instability is current density driven and close to the peeling limit. Figure reproduced from [128].*

Whilst the peeling mode is driven by the torque created by a finite current density at the plasma edge and no current in the vacuum region, the kink mode is driven by the derivative of the parallel current density.

In addition to the higher temperatures in the H-mode, the increase in edge current density is primarily due to the bootstrap current. Figure 4.12 presents an ELM cycle for the condition where the pressure gradient is low, the current density is high and the peeling mode is triggered. This could occur if the current diffusion time is short enough and the peeling boundary is crossed before the pressure gradient has reached the ballooning stability limit. Type III ELMs are regarded as peeling modes or alternatively as resistive peeling-ballooning instabilities. A possible type III ELM cycle, based on the peeling limit is shown in Figure 4.12.

- **Ballooning modes:** The ballooning modes are driven by the edge plasma pressure gradient and arise from the curvature of the tokamak geometry. The curvature effect is stabilizing on the high field side (HFS) and destabilizing on the low field side (LFS). The average of these effects is stabilizing for low pressure gradient but turns destabilizing if the pressure gradient becomes too high and drives ballooning modes that localize in the LFS region.

Figure 4.12 presents the situation where the pressure gradient is high, the current density is low and hence the ballooning mode is triggered. This scenario could occur if the plasma shaping is strong or there is a high edge density. This instability could be causing type II ELMs.

- **Coupled peeling-ballooning modes:** are driven by the steep edge pressure gradient and consequently a large edge bootstrap current [129] [130] [131]. It can be observed from Figure 4.12 that the stable window is limited both by edge current and pressure. Typically, the ballooning limit occurs at higher n numbers of the order 15-20 whereas the peeling limit is usually connected with lower $n \leq 5$. It can be interpreted that the limitation of pressure is due to the ballooning limit while the ELM crash itself occurs when peeling and ballooning modes couple in the upper right corner of the stability diagram.

The coupled peeling-ballooning model is the leading candidate for explaining the ELM onset as it is very successful in describing a number of experimental observations related to type I ELMs. However, it can certainly not be considered complete as it does not rigorously include several physics aspects such as sheared edge rotation, finite resistivity, two fluid effects etc. This however, does not lie within the scope of this thesis and will not be treated further.

4.4 ELM control

As described earlier, type-I ELMy H-mode is foreseen as the ITER baseline scenario. However, ELMs pose a serious concern in ITER because of the high transient heat and particle flux that can be deposited on the PFCs. At the same time, ELMs play a key role in regulating particle and impurity transport across the edge barrier and are thus instrumental in hindering the uncontrolled rise of density and impurity buildup which is observed in ELM-free H-modes. Therefore, reliable methods for the control of large type I ELMs are necessitated. Management of large type I ELMs can be excersized on four fronts:

- **Small/ELM-less regimes**: Operation in small or ELM-less regimes was described earlier in subsection 4.2.4. It can be noted that these regimes are in

general possible over a limited range of plasma parameters. Therefore, other active methods of ELM control are required.

- **Radiative dispersion**: Dispersing ELM energy loss by radiation before it reaches the divertor and PFCs. This is accomplished primarily via impurity gas seeding. On JET, the injection of Ar and N_2 has succeeded in reducing ELM energy by about 20% on the outer divertor and about 25 % on the inner divertor target. However, increase in the radiative power fraction above 65% causes a transition from type I ELMs to type III ELMs, leading to the so called radiating type III ELMy H-mode.

- **ELM triggering**: Destabilizing the plasma for triggering an ELM, before the stability limit is reached. Several techniques have been used for ELM triggering and include pellet injection [132][119], supersonic molecular beam injection (SMBI) [133][134], vertical kicks[135] and oscillating applied magnetic fields[136].

- **ELM suppression**: Controlling either the pedestal pressure gradient or the edge current density below the peeling-ballooning ELM stability limit. A prime example of this scheme is the control of ELMs by resonant magnetic perturbations (RMPs) [137][138].

ELM triggering provides additional transport by frequently triggering an instability over a small portion of the pedestal before the larger ELM engulfs the entire pedestal. A perturbation is introduced in the plasma and ELMs are triggered at a suitable frequency which is higher than the natural frequency. For this reason, such techniques are more widely known as ELM pacing. ELM pacing techniques rely on the observed inverse dependence of ELM size (W_{ELM}) on the ELM frequency (f_{ELM}) [139]:

$$W_{ELM} = 0.2 \left(\frac{W_{plasma}}{f_{ELM} \times \tau_E} \right), \tag{4.6}$$

where τ_E is the energy confinement time in plasmas with a stored energy W_{plasma}. However, the observed dependence of the effective ELM energy deposition area (A_{ELM}) on ELM size provides a caveat [140][141]. A broadening of A_{ELM} by a factor of up to 6 has been observed for large W_{ELM} ($\sim$ 10% of W_{plasma}) but is

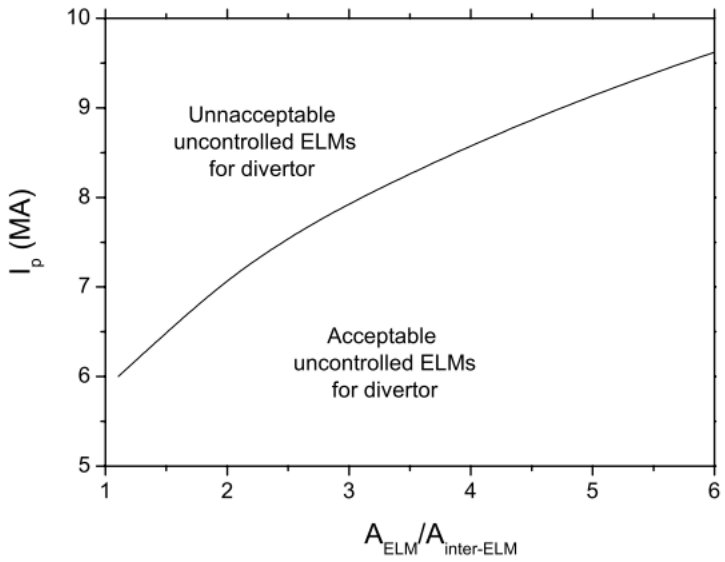


Figure 4.13: *Range of plasma current for ITER's H-mode operation for which uncontrolled ELMs are acceptable (from the point of view of divertor PFC erosion) versus broadening factor ELM energy deposition area with respect to the divertor power flux footprint between ELMs. Figure reproduced from [142]*

found to be much smaller for small W_{ELM}. This is likely to reduce the efficacy of ELM pacing techniques for the reduction of the peak ELM energy flux at the ITER divertor. However, ELM control via pacing methods finds applicability in conditions where uncontrolled ELMs would still exceed the limits posed by divertor erosion. Figure 4.13 presents the results reported in [143][142] where for a large degree of broadening in A_{ELM} (factor of 6), uncontrolled ELMs, in ITER, would lead to unacceptable divertor erosion for $I_p > 9.5MA$.

Furthermore, the recent ELM pacing experiments at DIII-D using lithium granules in contrast to frozen deuterium pellets, report on a reduction of the peak ELM heat flux at the outer strike point [144]. This result demonstrates the possibility of ELM pacing by non-fuel pellet injection which also has an added advantage of de-coupling ELM pacing from plasma fueling.

Finally, ELM pacing techniques will be needed in ITER for preventing an increase in the core concentration of impurities, in particular W, which will be produced at the divertor target [135][142]. JET operation with the ILW has shown that maintaining a sufficiently high f_{ELM} is critical for maintaining tolerable core W concentration [145]. The beneficial effects of a high f_{ELM} in limiting the concentration of high Z ions, has also been shown at AUG [146].

Chapter 5

Visualization of the operational space of ELMs

As introduced in chapter 2, information visualization aimed at facilitating human perception is an important tool for the interpretation of experiments on the basis of complex multidimensional data characterizing the operational space of fusion devices. Visualization of the machine operational spaces entails the representation of multidimensional diagnostic fusion data in a low-dimensional space, usually of two or three dimensions. This can provide physicists and machine operators with a convenient means and a useful tool for plasma monitoring and for studying data patterns (relationships, clusters) reflecting key regimes and their underlying physics.

In this work, two data visualization frameworks: *geodesic distance based multidimensional scaling* (GD-based MDS) and *geodesic distance based landmark muldimensional scaling* (GD-based LMDS) are developed and applied to the visualization of the operational space of ELMs and low and high plasma confinement regimes. It is noteworthy that the developed frameworks are general and can be applied to the study and analysis of various other plasma phenomena as well.

*The work presented here in chapter 5, has resulted in the following publications:

A.Shabbir, G.Verdoolaege, O.J.W.F.Kardaun, J.-M.Noterdaeme and JET-EFDA Contributors, "Visualization of the operational space of edge-localized modes through low-dimensional embedding of probability distributions," *Review of Scientific Instruments*, vol.85, issue 11, no.11E819, 2014.

A.Shabbir, G. Verdoolaege, G.Van Oost, J.-M.Noterdaeme, and JET-EFDA Contributors, "Visualization of tokamak operational spaces through the projection of data probability distributions," in *Proc. of* 40^{th} *EPS Conference on Plasma Physics*, P2.132, Espoo, Finland, July, 2013.

5.1 Visualization using GD-based MDS

GD-based MDS is applied for detecting cluster structure corresponding to type I and type III ELMs in an assembled dataset of JET plasma discharges. The presented methodology attempts to incorporate the substantial stochastic uncertainty affecting the ELM properties by representing the measurements of each discharge by a probability distribution. Through a probabilistic description of plasma signals, GD-based MDS aims to utilize the information content residing in the error bars (measurement error and statistical variation) associated with each measurement alongside being potentially capable of taking into account higher order statistical moments.

5.1.1 Plasma parameters describing the operational space of ELMs

The presented visualization framework is employed here for the visualization of clusters corresponding to type I and type III ELMs from a series of CW JET plasmas between the years 2000 and 2009 with MarkGB and MarkGBRS as divertors. In a second phase, this can also be used for the classification of ELM types. From the range of discharge numbers [50564, 76871], a selection of 69 JET plasmas pertaining to type I ELMs, 26 JET plasmas of type III ELMs and 5 JET plasmas [66105–66109] of the so-called type I high-frequency (HF) ELMs have been made. This constitutes the JET CW ELMy database (JET ELM DB-I) reported in appendix A.1: JET ELM-DBI and further used for analysis in chapter 7.

A threshold-based ELM detection algorithm has been developed (see chapter 7)

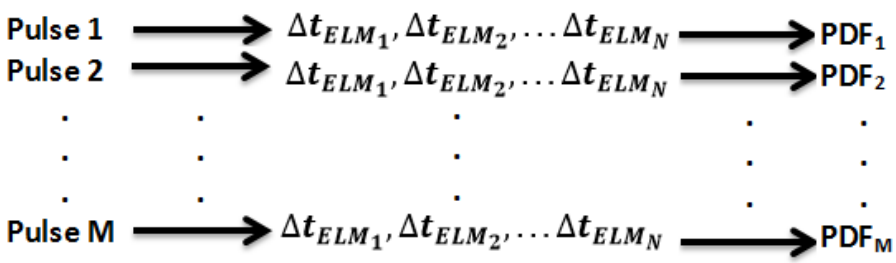


Figure 5.1: *From each plasma discharge with, let us say, $N+1$ ELM bursts, N waiting times are extracted which are then modeled by a suitable probability density function (PDF). M indicates the total number of discharges analyzed and the PDF used here is the Gaussian PDF.*

and employed for the extraction of inter-ELM time intervals (or waiting times) (Δt_{ELM}) from each plasma discharge using the D_α radiation signal from JET's inner divertor. Gaussian probability density functions (PDFs) are then used for capturing the statistics of inter-ELM time intervals, as shown in Figure 5.1.

In addition, density-averaged input power ($\langle P_n \rangle)(keV/s)$, normalized electron temperature ($\langle T_n \rangle)(keV)$, and line-integrated edge electron density (n_e) $[10^{19}\, m^{-2}]$ have also been included in the dataset. A probability distribution has been used for modeling the data corresponding to each plasma parameter. For simplicity it has been assumed that the error bars associated with each of the global plasma parameters pertain to a statistical uncertainty in the data, specifically that they represent a single standard deviation. Theoretically, the underlying probability distribution is Gaussian with mean the measurement itself and standard deviation the error bar. $\langle P_n \rangle$ and $\langle T_n \rangle$ are given as follows:

$$\langle P_n \rangle = \frac{1}{1.602 \times 10^3} \times \frac{P_{input}}{n_{e,19,(vol.avg)} \times Volume}\ keV/s, \tag{5.1}$$

where $P_{input} = P_{ohmic} + P_{NBI} + P_{ICRH}(W)$, and

$$kT = \langle T_n \rangle = \frac{1}{1.602 \times 10^3} \times \frac{1}{3} \times \frac{W_{thermal}}{n_{e,19,(vol.avg)} \times Volume}\ keV. \tag{5.2}$$

It is stressed that each plasma parameter is in fact a two dimensional quantity with the mean and standard deviation of the respective probability distribution constituting the two dimensions.

5.1.2 Visual maps

2D visual mappings for the operational space pertaining to type I and type III ELMs are presented in Figure 5.2, Figure 5.3 and Figure 5.4. In these maps, GD-based MDS is used for projecting the information. In Figure 5.2, inter-ELM time intervals, modeled with Gaussian PDFs, are projected and the ELM type for each discharge is indicated. The projection appears to be consistent with the information provided by the distribution of waiting times as the type I HF ELMs that have frequencies similar to type III ELMs are mapped in the proximity of type III ELMs. In addition, the type III ELM discharge 50567, which has a frequency ($f_{ELM} \approx 40Hz$) more typical

of type I ELMs, is projected close to the cluster of type I ELMs. Figure 5.3 explicitly tracks the change in ELM frequency and the consequent change in ELM type as one navigates through the map. Figure 5.3 is an indicator of the potential of these visual maps as they allow for tracking of changing values of a certain plasma parameter in the operational space. Figure 5.4(a) incorporates two additional parameters in the visual map, i.e. $\langle P_n \rangle$ and $\langle T_n \rangle$. The striking observation is that as more information is incorporated in the visualization, type I HF ELMs and type III discharge 50567 now lie with the clusters of type I and type III, respectively. Hence, incorporating additional global plasma parameters improves the accuracy of the map, with regards to the ELM type. Figure 5.4(b) incorporates n_e as another additional parameter, and more correct placement of type I HF ELMS and discharge 50567 can also be observed here. In addition, the line of best separation between type I and type III ELMs is shown in Figure 5.4(a) and Figure 5.4(b). This can aid in the classification of ELM behavior when the ELM type is unknown. Projection on this map of a new discharge, for which the ELM type is unknown, can provide an indication of the ELM type.

In Figure 5.4(a) and (b) cluster structure amongst the plasma discharges is identified. Further, the trends in plasma parameters along the clusters are also indicated. In both Figure 5.4(a) and (b), $\langle P_n \rangle$ and $\langle T_n \rangle$ increase from the bottom of the map to the top, as a transition is made from the clusters of type III ELMs to type I ELMs. Background color gradation indicates the broad trend of the changing

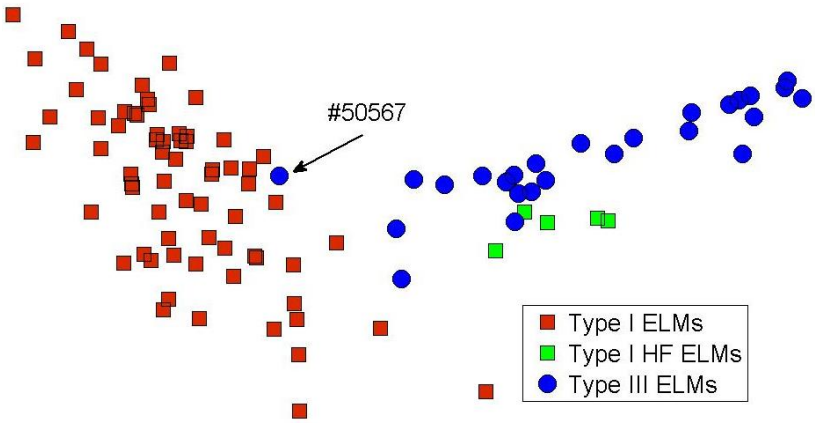


Figure 5.2: *Projection of inter-ELM time intervals using GD-based MDS. ELM types are indicated with Gaussian PDFs.*

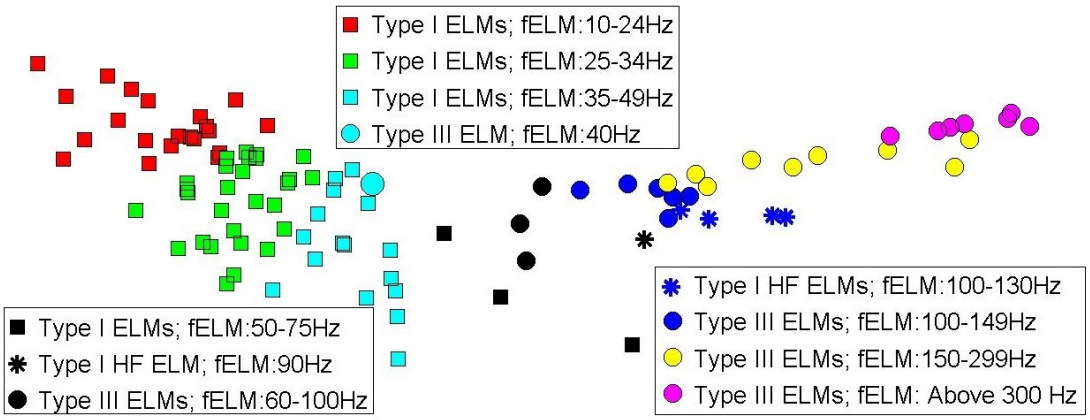


Figure 5.3: *Projection of inter-ELM time intervals using GD-based MDS. Different colors indicate ELM frequency ranges.*

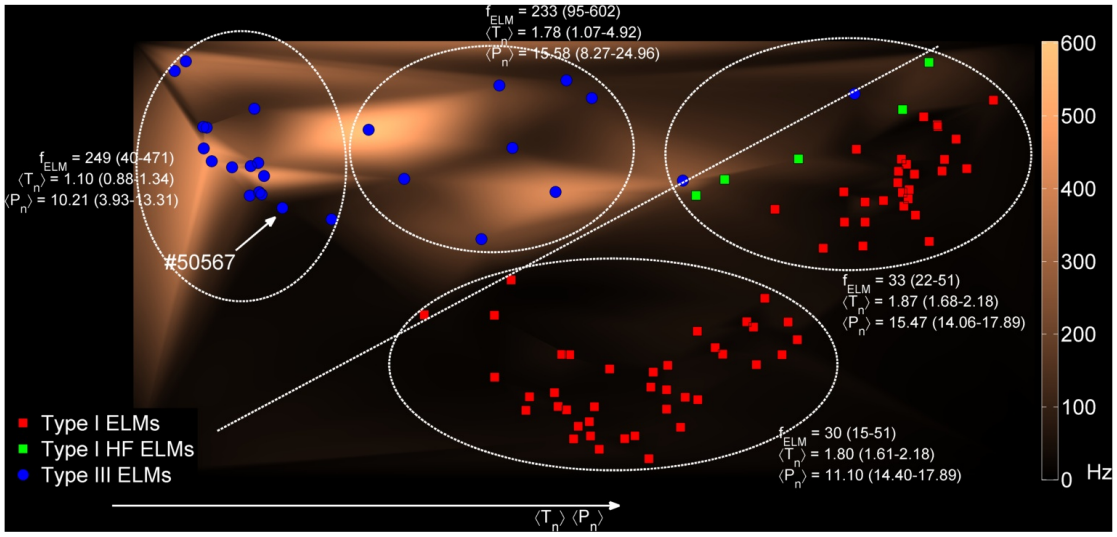
ELM frequency within the map. Finer structure in ELM frequency change can be further studied by using less flexible quadratic surfaces given by the regression analysis below.

Ordinary least squares linear regression using the two dimensions (x_1, x_2) provided by GD-based MDS as predictors for plasma parameters, yields appreciably good results. The results are given in Table 5.1. Each plasma parameter is successively regarded as the dependent variable. This provides an intuitive insight into the merit of the mappings, as despite the lack of a direct relationship between the dimension coordinates and each of the plasma parameters, the mappings prove to be reasonably reliable predictors of the plasma parameters. Goodness of the regression models is indicated by root mean square error (RMSE) and R-squared

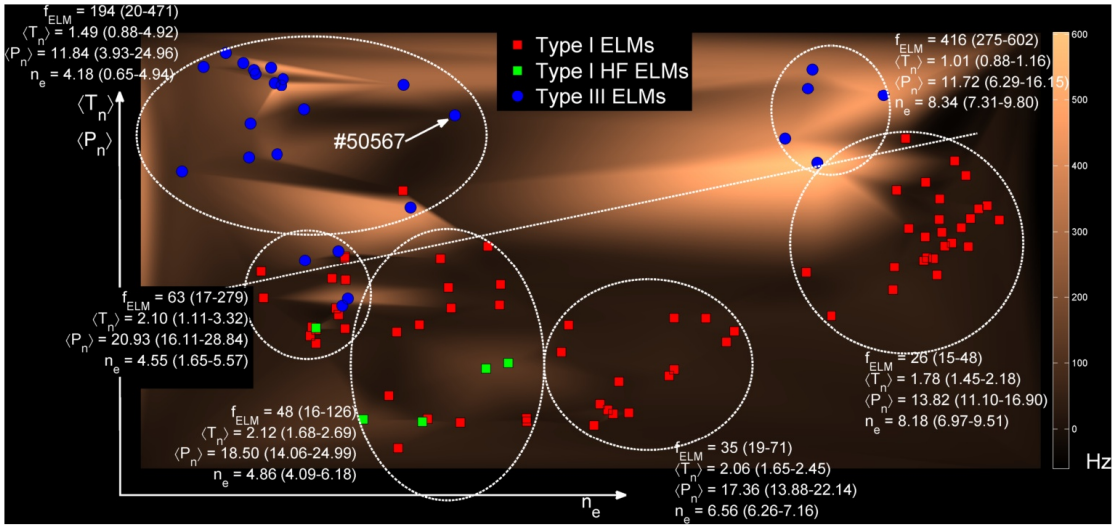
Model: $log\ y \sim 1 + x_1 + x_2 + x_1x_2 + x_1^2 + x_2^2$
$log\ n_e = 1.78 + 0.2x_1 - 0.003x_2 - 0.04x_1x_2 + 0.014x_1^2 + 0.031x_2^2$
$log\ T_n = 0.45 - 0.04x_1 + 0.28x_2 + (0)x_1x_2 + 0.12x_1^2 - 0.024x_2^2$
here, x_1 and x_2 are standardized co-ordinates of the two dimensions yielded by GD-based MDS

Dependent variable ($log\ y$)	R^2	RMSE
$\langle T_n \rangle$	0.76	0.14
$\langle P_n \rangle$	0.74	0.15
n_e	0.57	0.22
f_{ELM}	0.51	0.67

Table 5.1: *Regression for plasma parameters using the dimension coordinates (standardized) yielded by GD-based MDS as predictors.*



(a)



(b)

Figure 5.4: *Projections obtained using GD-based MDS. (a) A map of the distributions of ELM waiting times, $\langle P_n \rangle$ and $\langle T_n \rangle$. (b) A map of the distributions of ELM waiting times, $\langle P_n \rangle$, $\langle T_n \rangle$, and n_e. Clusters are identified and the mean value for each plasma parameter for the respective cluster is given. The range of the parameter values for discharges in each cluster is specified within brackets. The line of best separation between the type I and type III ELMs is also depicted. Furthermore, trends in plasma parameters across the map are highlighted. Background color gradation indicates the trend in ELM frequency.*

(R^2) presented in Table 5.1.

The presented visualization framework is generic, potentially capable of visualizing multi-machine data in a single map and can be adapted with relative ease if the plasma parameters are best described by non-Gaussian PDFs such as Weibull or lognormal. The framework can be used for investigating the influence of plasma parameters on ELM characteristics such as the effect of changing heating power on ELM type during a discharge. The developed tool can also potentially contribute to ELM control and mitigation, through visualization of, e.g., ELM frequency and size distributions, and by quantifying the influence on these distributions of various control parameters.

5.2 Visualization using GD-based LMDS

While MDS is a well-developed information visualization tool which yields a 2D map with minimal distortion of all pairwise distances between data points, it suffers from a polynomial computational and memory complexity and hence when the data size increases, it becomes too computationally expensive, for all practical purposes. As introduced in subsection 2.3.2, this problem is circumvented by LMDS, a computationally efficient adaptation of MDS which reduces the complexity from polynomial to linear.

The developed computationally efficient visualization framework of GD-based LMDS is deployed for the visualization of low and high confinement regimes from the International Tokamak Physics (ITPA) Global H-mode Confinement Database (ITPA database)[147]. The database contains more than 10,000 validated measurements of various plasma and engineering parameters, during discharges in 19 tokamaks.

5.2.1 Visual Maps

Our proposed GD-based LMDS framework regards each measurement as a sample from a Gaussian probability distribution with the mean the measurement itself and standard deviation the error bar. The visual map obtained via the application of GD-based LMDS is presented in Figure 5.5 (b). It can be observed that the mapping presented in Figure 5.5 (b) is only slightly less accurate when compared

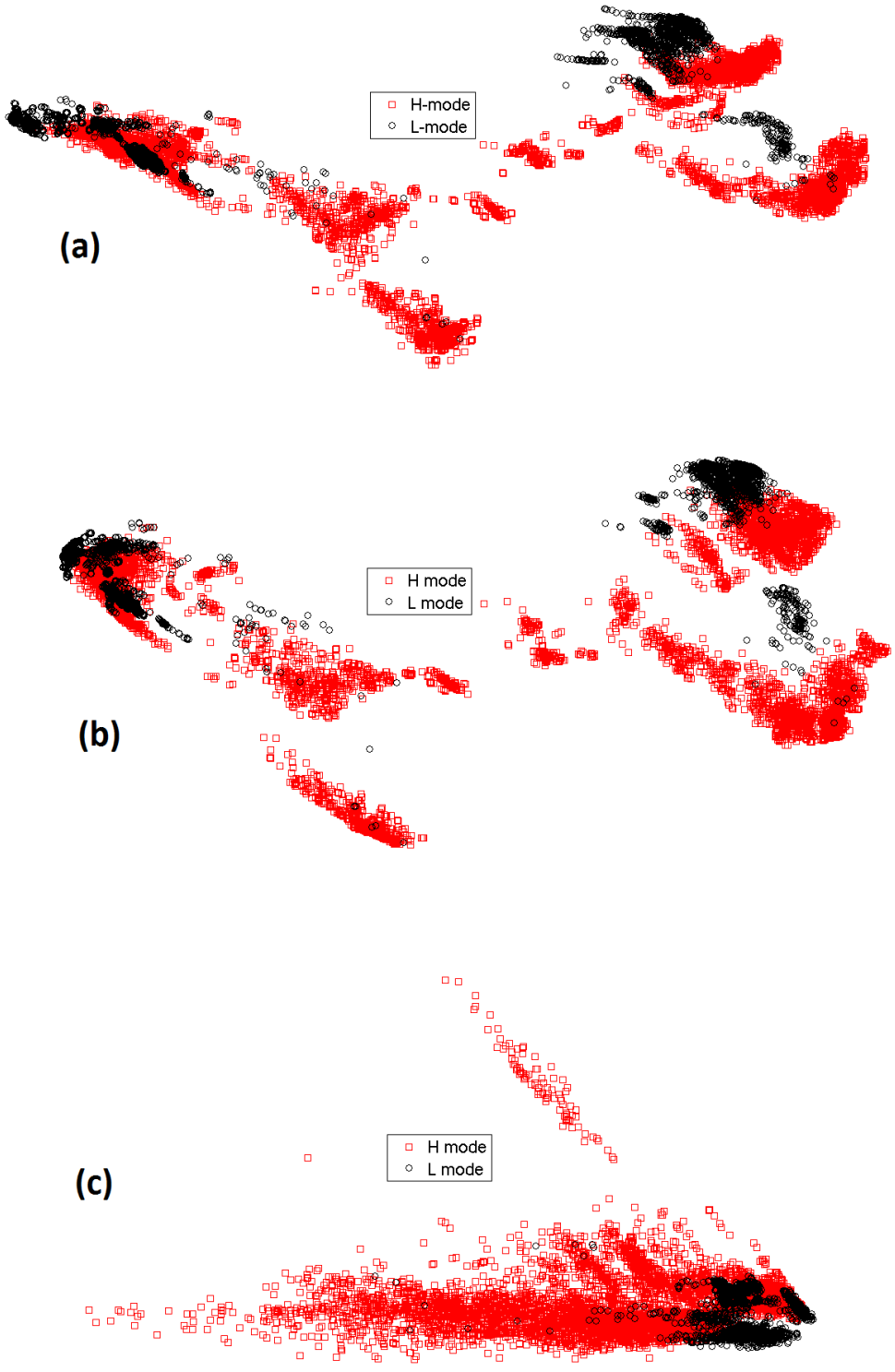


Figure 5.5: *Two-dimensional projections of the ITPA database, indicating L- and H-mode clusters. (a) Using GD-based MDS. (b) Using GD-based LMDS with 10 landmark points and (c) Ordinary LMDS (Euclidean distance based) and without accommodating measurement uncertainty.*

with the mapping obtained with GD-based MDS (Figure 5.5(a)), which is considered as the reference map. Meanwhile the execution time for obtaining the mapping by GD-based LMDS is reduced by a factor of 300 when compared to GD-based MDS.

Moreover, in order to further validate the attractive capabilities of our proposed GD-based LMDS framework, we compare the mapping with the map obtained in Figure 5.5 (c) through ordinary LMDS. Ordinary LMDS does not accommodate the measurement uncertainty by probabilistic modeling and operates by taking as input the matrix of Euclidean distances between measurements. It can be readily observed that in contrast to the mapping obtained by ordinary LMDS the mapping obtained by GD-based LMDS is much more informative and has little overlap between the classes. Application of GD-based LMDS yields a visualization which exhibits structure and provides a clear indication of the type of confinement regime.

Chapter 6

ELM regime classification by conformal prediction on an information manifold*

Aqsa Shabbir[1,2], *Member*, IEEE, Geert Verdoolaege[1,3], *Member*, IEEE, Jesús Vega[4] and Andrea Murari[5]

[1]Department of Applied Physics, Ghent University, Ghent 9000, Belgium
[2]Max-Planck Institute for Plasma Physics, Garching 85748, Germany
[3]Laboratory for Plasma Physics – Royal Military Academy (LPP – ERM/KMS), Brussels 1000 , Belgium
[4]Laboratorio Nacional de Fusión, CIEMAT, Madrid 28040, Spain
[5]Consorzio RFX, CNR, ENEA, INFN, Acciaierie Venete S.P.A, Università di Padova, Padua 35127. Italy

Abstract -Characterisation and control of plasma instabilities known as edge-localised modes (ELMs) is crucial for the operation of fusion reactors. Recently, machine learning methods have demonstrated good potential in making useful inferences from stochastic fusion data sets. However, traditional classification methods do not offer an inherent estimate of the goodness of their prediction. In this work, a distance-based conformal predictor classifier integrated with a geometric-probabilistic framework is presented. A first benefit of the approach lies in its comprehensive

*The work presented here in chapter 6, has been published in this form as:
A.Shabbir, G.Verdoolaege, J.Vega and A.Murari, "ELM classification by conformal prediction on an information manifold," *IEEE Transactions on Plasma Science*, vol.43, no.12, pp.4190-4199, 2015.

treatment of highly stochastic fusion data sets, by modeling the measurements with probability distributions in a metric space. This enables calculation of a natural distance measure between probability distributions: the Rao geodesic distance. Secondly, the predictions are accompanied with estimates of their accuracy and reliability. The method is applied to the classification of regimes characterized by different types of edge-localized modes based on measurements of global parameters and their error bars. This yields promising success rates and the estimates of goodness of the predictions increase the confidence of classification by ELM experts, while allowing more reliable decisions regarding plasma control and at the same time increasing the robustness of the control system.

Index terms **-Conformal predictors, edge-localized modes, geodesic distance, information manifold.**

6.1 Introduction

High confinement or H-mode plasmas in tokamaks are usually characterized by cyclic instabilities near the plasma edge, referred to as *edge-localized modes* or ELMs. ELMs result in a sudden exhaust of particles and energy but are nonetheless advantageous for attaining stationary plasma conditions as they result in impurity and helium ash expulsion. With ELMs as the basis for distinction, H-mode plasmas can be roughly categorized into three types: ELM-free H-mode, H-mode with small ELMs and H-mode with large or type I ELMs. The H-mode with relatively large low-frequency type I ELMs has become the reference plasma scenario for ITER, and beyond. Characteristics of ELMs, their control and comprehensive physical understanding are crucial for ITER and next step fusion devices. On one hand the beneficial properties of ELMs, in terms of enhanced edge particle transport are well recognized, on the other hand, there has been concern that on future large devices giant ELM bursts could damage divertor and first wall surfaces or disrupt internal transport barriers [42]. This has motivated intensive research for accomplishing effective ELM control and mitigation. Optimization of control and mitigation mechanisms and enhancement of the physical understanding

necessitates the discrimination of different observed classes of ELMs. In contrast to the existing mostly phenomenological categorizations of ELM types, this work is aimed at developing a data-driven methodology for automatic classification and discrimination of ELMs.

Recently, machine learning and pattern recognition techniques have shown substantial potential in data-driven studies of fusion plasmas by extracting useful patterns of interest from fusion data [148] [62] [89] [149]. This yields an important tool for real-time plasma control, e.g. in ITER, in order to maintain good plasma equilibrium or control certain types of instabilities. Moreover, a data-driven study of the primary physical variables that determine the confinement regimes and instabilities, such as ELMs, can improve substantially the understanding of the governing physical mechanisms.

The objective of the present work is twofold. First, we wish to contribute to the discrimination of diverse ELM behavior by presenting an effective methodology for quantitative distinction between ELM types. Second, for practical purposes we aim to contribute to the dependability and robustness of control strategies by providing a discriminator for ELM types equipped with estimates of reliability and accurateness. We present an automated classification system for ELM types and apply our method to classify regimes with small and type I ELMs. The system, currently, makes use of a standard set of global plasma and engineering variables related to plasma confinement. The focus of the current work is to obtain better classification rates compared to existing classifiers and thus the obtained success rates can be further optimized by using more informative plasma and engineering parameters.

The act of classification is fundamentally related to the occurrence of *clustering structure* in the data space, where each cluster of measurement points corresponds to a certain plasma phenomenon, such as a specific type of ELMs. Hence our method falls within the domain of *pattern recognition methods*, with the clusters constituting a pattern in the data space, reflecting an important aspect of the physics of the plasma.

The classification system proposed in this paper is integrated with the probabilistic data representation framework presented earlier by Verdoolaege *et al.* [62][89]. The primary motivation for this framework is the substantial uncertainty

that frequently characterizes the measurements of plasma quantities, which may contain both a stochastic and systematic component. The main factors contributing to stochastic uncertainty are hardware noise and plasma fluctuations, since these can usually not be modelled tractably in a deterministic way. The proposed framework takes into account the statistical error bars or, more generally, the stochastic features of the data, by modelling the data with suitable probability distributions. In order to characterize the data patterns, such as clusters, in the associated probabilistic space, a similarity measure between probability distributions is required. The mathematical field of information geometry provides an appropriate similarity measure between probability density functions (PDFs), which are interpreted as points on a Riemannian differentiable manifold, or *information manifold* [33][55][48]. The PDF parameters provide a coordinate system on the manifold and the Rao geodesic distance (GD) serves as a natural similarity measure between PDFs. The classifier, which then operates in this information space, is based on *conformal predictors* (CPs), first described by Vovk *et al.*[150] and Saunders *et al.* [10]. Conformal prediction offers various advantages over the traditional machine learning methods (MLMs). Most noteworthy, they provide information about their own accuracy and reliability with the only assumption of randomness of the data samples. Also known as the iid hypothesis, the randomness assumption implies that all training samples are independent of each other and are identically distributed according to the same (but unknown) distribution [151]. Unlike traditional MLMs, CPs do not enforce a rigid separation between learning and prediction, but learn dynamically alongside making predictions. Furthermore, they do not require prior probabilities as the Bayes classifier and also have the ability to detect ambiguities in the classification task, i.e., when a unique class cannot be assigned to a new example.

In this paper a computationally efficient nearest-neighbor CP coupled with the geometric-probabilistic data representation framework is deployed for classification of H-mode plasma regimes into H-mode with small ELMs and H-mode with type I ELMs. The proposed technique is compared with a discriminant analysis classifier and a nearest-neighbor classifier, which are well-established state-of-the-art MLMs. The presented technique not only yields higher classification accuracy, but also returns a quantitative estimate of the prediction's accuracy and reliability, which

traditional MLMs do not provide. Furthermore, the classification performance is calculated for both the geodesic distance geometry of the data and the conventional Euclidean distance. The geodesic distance improves the classification performance, establishing itself as a natural similarity measure between probability distributions lying on an information manifold.

The outline of the paper is as follows. In section 6.2 we discuss the modalities of our proposed geometric-probabilistic framework and the details of the approach. section 6.3 discusses the application of conformal predictors to ELM identification in relation to our modeling framework. section 6.4 presents the experimental setup, visualization and classification results and their analysis. section 6.5 concludes the paper.

6.2 A geometric-probabilistic pattern recognition framework

6.2.1 The geometry of probability distributions

The Fisher information can be regarded as a metric tensor (Fisher-Rao metric) on an information manifold, which is a Riemannian differentiable manifold formed by a family of PDFs, such as the Gaussian family [62][89]. Once the metric is known, geodesic equations can be established and solved, allowing for the calculation of the geodesic (shortest-path) distances on the manifold [55][48]. Given a probability model $p(x|\theta)$ for a vector-valued variable x, labelled by an m dimensional parameter vecttor θ, the components of the Fisher information matrix $g_{\mu v}$ are defined through the relations

$$g_{\mu v}(\theta) = \mathbb{E}\left[\frac{\partial^2}{\partial\theta^{\mu}\partial\theta^{v}} lnp(x|\theta)\right], \tag{6.1}$$

$$\mu, v = 1, ...m.$$

6.2.2 The geometry of the univariate Gaussian distribution

In this paper we model the data using a simple univariate Gaussian model. The Fisher-Rao metric for the Gaussian distribution, parameterized by its mean μ and

standard deviation σ, can be given via the quadratic line element [56]:

$$ds^2 = \frac{1}{\sigma^2} d\mu^2 + \frac{2}{\sigma^2} d\sigma^2. \tag{6.2}$$

A closed-form expression exists for the GD, permitting a fast evaluation. Indeed, for two univariate Gaussian distributions $p_1(x|\mu_1, \sigma_1)$ and $p_1(x|\mu_2, \sigma_2)$, parameterized by their mean μ_i and standard deviation $\sigma_i (i = 1, 2)$, the GD is given by [56]

$$GD(p_1||p_2) = \sqrt{2} ln \frac{1+\delta}{1-\delta}, \tag{6.3}$$

$$\delta \equiv \left[\frac{(\mu_1 - \mu_2)^2 + 2(\sigma_1 - \sigma_2)^2}{(\mu_1 - \mu_2)^2 + 2(\sigma_1 + \sigma_2)^2} \right]^{\frac{1}{2}}. \tag{6.4}$$

A convenient Gaussian geometric model is provided by the Poincaré half-plane, which is represented in Figure 6.1(a). The horizontal axis corresponds to the mean μ of the Gaussian distribution, while on the positive part of the vertical axis the standard deviation σ is represented. Every point in this half-plane corresponds to a unique Gaussian and the geodesics between two points are half-circles as well as half-lines ending on the horizontal axis, the latter connecting distributions that differ only in their standard deviation (not drawn). The distance between points along one of these curves in the Poincaré half-plane is the same as the actual geodesic distance between the points. The evolution of the distribution along an example geodesic is shown in Figure 6.1(b).

Finally, in the case of multiple independent Gaussian variables it is easy to prove that the squared GD between two sets of products of distributions is given by the sum of the squared GDs between corresponding individual distributions [56].

6.3 Conformal predictors

In classification systems, each observation (or sample) is expressed as an ordered pair (x_i, y_i) where x_i is a feature vector (i.e. the set of parameters that characterize the sample i) and y_i is the class label of observation i, where the set of labels is finite and usually small. Given a data set of N samples, a conventional MLM uses a subset of the data set, which it designates as a training set for determining the prediction rule. Then follows a testing phase, wherein a subset of the dataset is

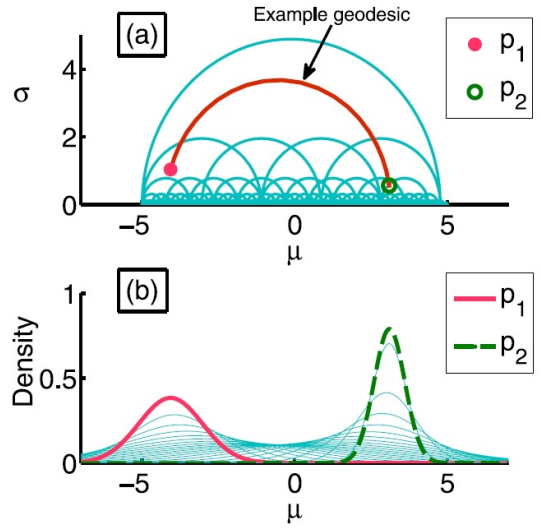


Figure 6.1: *(a) Illustration of the Poincaré half-plane with several half-circle geodesics, one of them between points p1 and p2. (b) Probability densities corresponding to points p_1 and p_2 indicated in (a). The densities associated with some intermediate points on the geodesic between p_1 and p_2 are also drawn.*

used for determining the goodness of the prediction rule. Conformal predictors obviate the need of a distinct training and testing phase, which is the premise of classical machine learning methodologies. CPs offer so-called blended learning and prediction, as they learn and predict at the same time, continuously improving their performance as they carry out each prediction and discover how accurate the prediction was. Samples that get classified are added to a hypothetical "bag" of samples and participate in the classification of the next incoming samples.

CPs estimate the goodness of their prediction by means of two figures of merit: *confidence* and *credibility*. Confidence gauges the reliability of the prediction, while credibility is an indicator of how representative the training set is for the new sample that is to be classified. New confidence values are obtained at each classification, taking into account both the previous samples that have been classified and all possible labels for the current one. For classifying each incoming sample, CPs evaluate how different the current sample is from each cluster (class) within the bag samples by determining a "nonconformity score" for the current sample with respect to each cluster (class) within the bag samples. In this work, a nearest-neighbor scheme is used for determining the nonconformity score. Essentially, the nonconformity score for the current sample is provided by its distance to its nearest neighbors for both classes, amongst the bag samples. Specifically, the

nonconformity α_i of a given sample i is calculated as

$$\alpha_i = \frac{d_{i-SL}}{d_{i-DL}}, \tag{6.5}$$

where d_{i-SL} is the distance to sample i's nearest neighbor in the bag with the same label and d_{i-DL} is the distance of sample i's nearest neighbor in the bag with a different label.

The nonconformity score for sample i is computed with respect to both classes, assuming membership of sample i of each of the classes $j = 1, 2$ in turn. By doing this for each sample, a ranking can be determined of the nonconformity scores. Then, for each class j a p-value is is calculated based on this ranking, namely:

$$p_j = \frac{\#\{i = 1, ..., M | \alpha_i \geq \alpha_q\}}{M}. \tag{6.6}$$

Here, M is the number of bag samples, i.e. the samples that have already been classified, and α_q is the nonconformity score for the sample that is to be currently classified. The p-value is essentially the fraction of bag samples that are at least as different as the current sample. The current sample is assigned to the class with the largest corresponding p-value. The largest p-value itself is referred to as the credibility, while the complement of the other p-value is the confidence of the classification task:

$$Credibility = max(p_j),\ j = 1, 2 \tag{6.7}$$

$$Confidence = 1 - min(p_j),\ j = 1, 2 \tag{6.8}$$

The smaller p-value is essentially the probability of the prediction being in error and thus the probability of correctness of the current prediction is automatically quantified by the confidence.

In case where the p-value assigned to each class is the same, CP deems that a unique class cannot be assigned to the current sample. It refrains from making any (possibly incorrect) decision and separates the current sample whilst labeling it ambiguous.

Despite the numerous advantages offered by CPs, the method can become computationally expensive and thus infeasible in real time, for very large data sets. This is a direct consequence of the dynamic learning capabilities possessed by CP.

This limitation is overcome by deploying a computationally efficient variant of CP: inductive conformal predictors (ICPs) [151]. ICPs offer a compromise between dynamic learning and computational time, without causing degradation of the classification performance. ICPs divide the dataset into two sets: the *proper training set* and the *calibration* set. The proper training set, similar to a conventional MLM, is used for computing the decision rule once, which is dynamically improved as each sample from the calibration set (a pseudo test set) is classified. In this work, ICPs as well as a theoretically pure form of CPs known as transductive conformal predictors (TCPs) are each applied. TCPs in contrast to ICPs require a minimalistic proper training data set and at the least one sample per class suffices.

As a similarity measure in calculating the nonconformity scores, we first considered the Euclidean distance between the sample's feature vector and that of the bag samples. Then we compared its performance to that of the GD, this time treating the features as quantities with an error bar, hence Gaussian distributions.

6.4 ELM identification

6.4.1 Physics picture of ELM types

The physical mechanisms of the different observed classes of ELMs are complex. As a result, no unified first principles theory describing ELMs exists. Type I ELMs mainly seem to be driven by the steep pressure gradient, whereas small ELMs appear to be controlled by the absolute value of the edge temperature along with steep pressure gradient. This suggests that small ELMs are linked to resistive MHD phenomena whereas type I ELMs are associated with ideal MHD [20][19][152]. Considerable progress has been made in ELM modeling activity [153][154], with the peeling-ballooning model appearing as the leading candidate for explaining the trigger for the ELMs. This model builds on the two instability sources near the plasma edge namely, current and pressure gradients. It outlines a pseudo-triangular operating diagram for ELMs in the space of the ballooning pressure gradient and the normalized edge current.

Currently, type I ELMs and small ELMs are primarily distinguished by their response to increased heating power. The ELM repetition frequency for type I

ELMs increases with increasing power and decreases for small ELMs.

An alternate way of distinguishing between the two classes is to compare temperatures and densities at the pedestal top. In [155] it is shown that type I ELMs are clustered around a hyperbola of constant, high pedestal pressure. This constant corresponds to the theoretically predicted onset of pressure driven, ideal MHD ballooning mode instability. Small ELMs appear to occur below a critical pedestal temperature $T_{e,crit}$ which tends to increase with the toroidal magnetic field. Further, on the $n_{edge} - T_{edge}$ diagram [155] they are seen as two clusters: one at low $T_{e,ped}$ and high n_{ped} and the other at high $T_{e,ped}$ and low n_{ped}.

A crucial distinction between type I and small ELMs is their effect on plasma confinement. Type I ELMy H-modes have superior overall plasma confinement but the ELM size possess serious concerns for future fusion machines. On the contrary, the size of small ELMs offers no concern for the machine operation but the energy confinement time is 10-30% below that in type I H-mode [19][152].

In order to predict ELM behavior in next step fusion devices and ensure operation in the desirable ELMy regimes, development of an automated discrimination scheme for ELMs is required, constituting the starting point of this work.

6.4.2 ITPA database

In this work for ELM regime identification we employed measurements from the International Tokamak Physics Activity (ITPA) Global H-mode Confinement Database (DB3, version 13f), henceforth referred to as the "ITPA database" [147][156]. The ITPA database contains more than 10,000 validated measurements of various global plasma and engineering variables at one or several time instants during discharges in 19 tokamaks. The data have been used extensively for determining scaling laws for the energy confinement time, mainly as a function of a set of eight plasma and engineering parameters: plasma current, vacuum toroidal magnetic field, total power loss from the plasma, central line-averaged electron density, plasma major radius, plasma minor radius, elongation and effective atomic mass. We have used the same eight global variables to discriminate between type I and small ELMs. Specifically, all database entries with a confinement mode labeled HGELM and HGELMH were considered to belong to the H-mode region with type I ELMs and all database entries tagged HSELM and HSELMH were regarded as belonging to the

Machine	Total samples	Type I ELMs	Small ELMs
ASDEX	445	287	158
AUG	583	498	85
CMOD	46	0	46
COMPASS	26	13	13
DIII-D	343	249	94
JET	1780	980	800
JFT-2M	76	0	76
JT60-U	89	35	54
PBXM	80	19	61
PDX	117	48	69
TCV	15	15	0
TFTR	99	5	94
TDEV	10	0	10
START	9	0	9
MAST	12	0	12
NSTX	6	0	6

Table 6.1: *Total number of samples from each tokamak in the ITPA database belonging to the H-mode region with ELMs. The number of samples per class, i.e. small and Type I ELMs, is also given.*

H-mode region with small ELMs. For current work, the database entries have been normalized to bring all variables in proportion with one another prior to subsequent operations.

It should be noted that classification of ELM characteristics based on global non-time-resolved data is a considerable challenge. Indeed, in addition to the information contained in the global time-averaged values of the plasma parameters, space-resolved measurements, near the plasma boundary, of the plasma density and temperature could easily improve the recognition rates. Similarly, estimates of changes in the thermal and fast particle energy content per ELM burst and measurements of ELM frequency obtained from time traces of plasma quantities, such as the D_{alpha} radiation, can also considerably improve the predictive capacity of the method. However, in the present work we did not yet take into account these additional sources of information, although our method is perfectly able to incorporate and treat these data.

The ITPA database lists typical error estimates of measurements for the various plasma and engineering variables. This represents very limited information on the probability distribution underlying each quantity. Nevertheless, effective utilization of this limited information proves beneficial. In this work it is assumed that the error

bars pertain to a statistical uncertainty in the data, specifically that they represent a single standard deviation. According to the principle of maximum entropy, the underlying probability distribution is Gaussian with mean the measurement itself and standard deviation the error bar. Also, it is supposed that for stationary plasma conditions, all variables are statistically independent and so the joint distribution factorizes. This means that the joint distribution for the eight variables mentioned above is assumed to be just the product of the individual univariate Gaussian distributions. Clearly, this is a strong assumption and it is imposed here mainly for keeping the calculations tractable. It is noteworthy that our formalism has no difficulties with the heterogeneous sources of the measurements, coming from different tokamaks and possibly with different error bars for essentially the same quantities. The reason is that the error estimates are automatically embedded in the probabilistic data description.

The number of samples from each tokamak belonging to the H-mode class with ELMs is given in Table 6.1. Further, the numbers of small and type I ELM samples per machine are also listed.

6.4.3 Visualization

Visualization of high-dimensional data sets through a projection in the two-dimensional Euclidean plane is a useful tool for enabling plasma physicists to gain knowledge about the internal structure of the data and relationships in it. Its goal is to amplify human cognition and provide an intuitive insight into the possible interactions and relationships in complex and frequently large data sets [157]. Hence, visualization of the data within the region of the operational space corresponding to H-mode with ELMs can be very useful because it can potentially yield enhanced insight in the configuration of the operational space. It can convey important information regarding the conditions, under which specific plasma regimes occur, as well as the "distance" of the current plasma conditions from a certain desired or undesirable regime. Visualization of the operational space is not a straightforward task as the information is not normally directly available, since the number of variables labeling the operational space is often greater than two. Hence the dimensionality of the data space is higher than two, preventing a simple plot of the data in a two-dimensional diagram. Moreover, in our framework each

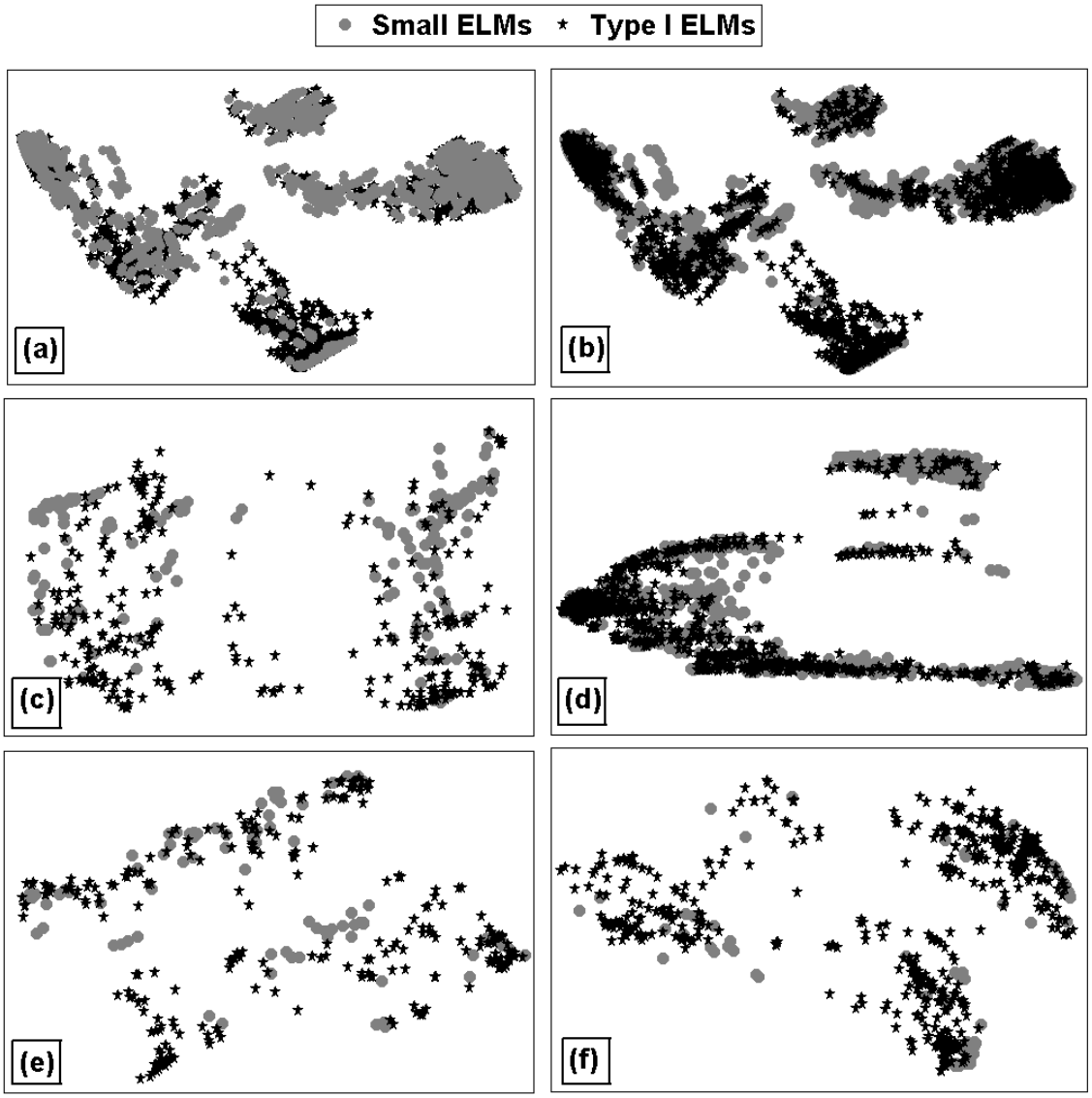


Figure 6.2: *2-D projections using MDS with indicated small and type-I ELM clusters. (a) ELM data from the entire ITPA database with small ELMs on top. (b) ELM data from the entire ITPA database with type-I ELMs on top. (c) ELM data from ASDEX. (d) ELM data from JET. (e) ELM data from DIII-D. (f) ELM data from AUG.*

measurement is represented by a Gaussian probability distribution with a mean and an error bar. This distribution cannot be represented by a point in a Euclidean space but naturally lies on a curved Riemannian manifold. Therefore, data visualization is a natural starting point in distinguishing between regions of different ELMs, which essentially are found in neighboring or overlapping regions of the operational space.

In this work, visualization of the high-dimensional and/or probabilistic (non-Euclidean) data is obtained by projection of the data onto a two-dimensional Euclidean plane. To do this, we use metric multidimensional scaling (MDS), which is a well-regarded information visualization technique [76] and is widely used

in perceptual mapping. MDS provides a two-dimensional mapping of the ITPA database, which originally spans eight dimensions (16 in case the standard deviation of each measurement is counted as an extra parameter). In order to calculate the distance in the original high-dimensional data space, we use the GD in the case when the probabilistic representation of the data is taken into account.

A projection using MDS is shown in Figure 6.2 for the entire ITPA database. Certainly, the visual map is an approximation of the original configuration, but nevertheless MDS yields a projection of points in the Euclidean space with least distortion of all pairwise distances; i.e. the mapping is approximately isometric. Hence, the real value of the projection lies in the relative position of the points with respect to each other and in contrast to usual scientific visualizations the coordinate axes are less significant. Further, a visual map from a subset of the data is also plotted in Figure 6.2. These are ASDEX, AUG, JET and DIII-D, as these machines are the major contributors to the ELM date in the ITPA database. Visualizations in Fig. 2 incorporate the measurement uncertainty as MDS uses the GD between Gaussian product distributions. It can be readily noted from Figure 6.2 that there is a considerable overlap between the ELM classes, rendering the classification task a veritable challenge (although it should always be remembered that the visualization is a projection, inevitably resulting in information loss). Further it can be seen in Figure 6.2(c) and (d) that the data of ASDEX and JET roughly conform to two clusters. This distribution is due to different levels of plasma current and toroidal fields in the machines. Also, it can be observed from visual inspection of Figure 6.2 (f) that the data of AUG is heavily unbalanced with very few samples from the H-mode region with small ELMs. Despite these constraints imposed by the data set, our classification scheme is able to attain a relatively good separation between the two classes.

6.4.4 Classification via conformal prediction

The experiments were performed for 20%, 50% and 70% of the total data being treated as a proper training set, followed by the dynamic learning for the remaining data. Proper training data were selected at random from the entire database, while ensuring the same balance with respect to the class variable as was present in the original data, i.e. if, say, the original data contained 70% samples from class 1 and

Initializing training data (%)	SR (%)	ER (%)	AM (%)	CO (%)	CR (%)
0.01	75.98	23.89	0.108	92.3	57.7
20	77.39	22.57	0.034	92.5	55.4
50	78.85	21.06	0.108	92.7	49.6
70	78.58	21.31	0.089	93	45.6

Table 6.2: *Average success rates (SR) (%), error rates (ER) (%), ambiguities (AM) (%), average confidence (CO) (%) and average credibility (CR) (%) for the classification of ELM types by transductive conformal predictor (TCP) using various sizes of the proper training data set and with the GD as the similarity measure.*

Initializing training data (%)	SR (%)	ER (%)	AM (%)	CO (%)	CR (%)
0.01	72.85	27.04	0.11	90.5	56.8
20	72.8	27.2	0	89.6	55.7
50	72.84	27.16	0	89.6	51.3
70	73.01	26.99	0	89.8	48.1

Table 6.3: *Similar to Table 6.2, but with the Euclidean distance as the similarity measure.*

30% samples belonged to class 2, then the samples which constitute the training set also maintain the same ratio with respect to class label. The results are given in Table 6.2. Transductive conformal prediction is also carried out for reference. This is shown in Table 6.2 as the entry which uses 0.01% of the total data as the initial proper training data. The GD is the distance measure of choice used to calculate the nonconformity scores, for the results presented in Table 6.2. Similar experiments were conducted using the Euclidean distance, operating on the measurement values without consideration of the error bars. These results are presented in Table 6.3. Table 6.2 and Table 6.3 each report the success rate (SR) for classification, which is the average over the two classes for correct predictions made as a percentage of total predictions. Similarly, the error rate (ER) is provided, which is the average over the two classes for the incorrect predictions made as a percentage of total predictions. Also listed is the ambiguity (AM), i.e. the ratio of the number of samples for which a prediction could not be made, for the total number of samples in the data set. The last two columns for each table contain the average values of confidence and credibility for the predictions made. The SRs achieved with the GD and Euclidean similarity measures are also illustrated in Figure 6.3.

The most noteworthy outcome is that the GD gives a superior performance in contrast to the Euclidean distance, both in terms of success rate and the average

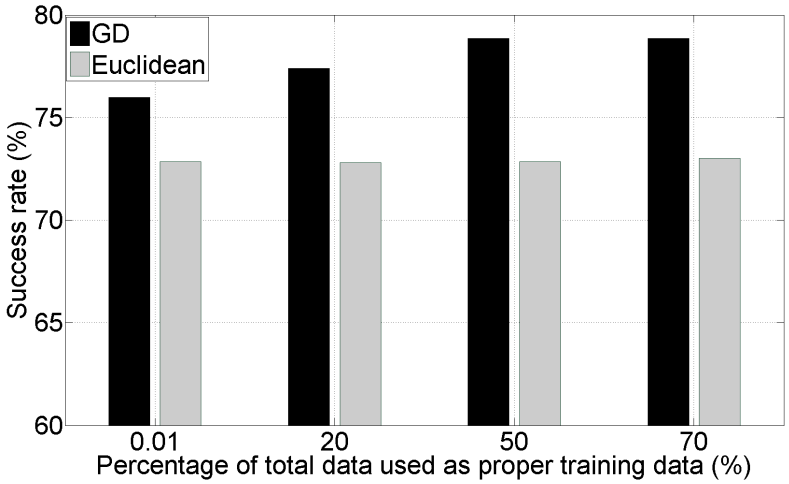


Figure 6.3: *SRs (%) for CP.*

confidence level of the predictions. This establishes that exploiting the information content residing in measurement uncertainty is important for identifying ELM types.

Furthermore, the CPU time (in seconds) for the CP classification obeys a linear law:

$$t = 1.30n + 48, \tag{6.9}$$

where n is the number of samples which constitute the calibration set, i.e. are dynamically classified. ICPs provide a significant reduction in computational time, as they effectively reduce the size of the calibration set without introducing degradation in success rates.

6.4.5 Comparative analysis

In previous works, discriminant analysis has been used for ELM identification [158]. To allow a homogeneous comparison between CPs and other well-established MLMs we perform the classification of H-mode with small and type I ELMs using linear discriminant analysis (LDA), quadratic discriminant analysis (QDA) and a 1-nearest-neighbor classifier. 50% of the data is used as training data, thus enabling a direct comparison with the entry with 50% of the data used as proper training data in Table 6.2 and Table 6.3. Balance with respect to the class variable is kept intact in the randomly selected training data. Each experiment is repeated 10 times with different random training sets and thus each mentioned result is in fact an average over 10 replications. The average SRs for classification alongside their standard deviation (STD) are given in Table 6.4 .

Classifier		SR (%)	STD (%)
LDA		60.26	0.39
QDA		68.65	0.33
1-NN	GD	70.50	0.97
1-NN	Euclidean	67.47	0.48

Table 6.4: *SR (%) and the corresponding standard deviations (STD) (%) for the classification of regimes with different ELM types based on linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), 1-nearest neighbor (1-NN) using GD and 1-NN using the Euclidean distance.*

It can be readily seen that CPs provide a significantly higher success rate in contrast to well-established MLMs, in addition to providing an estimate of the classification accuracy (confidence) and reliability (credibility).

6.4.6 Validation

We further empirically validated the performance of our classification scheme and the obtained results using N-fold stratified cross-validation. This is an established model validation technique, since for an optimal choice of N, it reduces the bias in the prediction output while combating variance and yet being computationally feasible. The mechanism behind cross-validation is illustrated in Figure 6.4. It operates by dividing the available data into roughly N equal parts and then iteratively training and testing the classification scheme using N-1 parts for training and the remaining one part for testing. Hence each sample in the data set gets eventually used for training and testing. We performed all our experiments using 10-fold cross-validation. The obtained results are given in Table 6.5 and Table 6.6. The success rates are higher for each classification scheme, compared to the results in the previous section. However, the important observation is that CP consistently

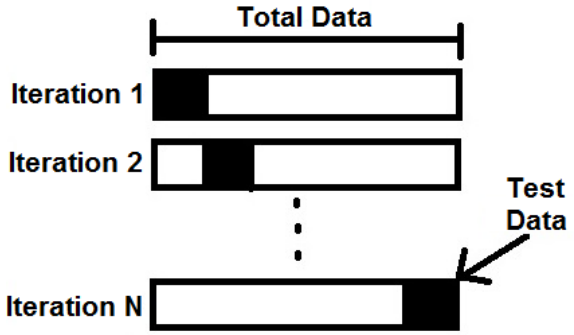


Figure 6.4: *Prototype for N-fold cross validation.*

Classifier	Distance Measure	SR (%)	ER (%)	AM (%)	CO (%)	CR (%)
CP (ICP)	GD	80.19	19.69	0.134	94.1	43.5
	Euclidean	73.5	26.48	0.027	90.2	48.1

Table 6.5: *SR (%), ER (%), AM (%), CO (%) and CR (%) for the classification of ELMy regimes with 10-fold cross-validation using inductive conformal predictor (ICP) with the GD and the Euclidean distance as similarity measures.*

Classifier		SR (%)	STD (%)
LDA		59.89	0.92
QDA		68.28	0.81
1-NN	GD	78.27	0.88
1-NN	Euclidean	72.85	0.67

Table 6.6: *SR (%) and the corresponding STD (%) for the classification of ELMy regimes with 10-fold cross-validation using LDA, QDA, 1-NN using the Euclidean distance and 1-NN using the GD.*

performs better than the other techniques. Furthermore, again the GD measure gives the better performance as compared to the Euclidean distance. Figure 6.5 illustrates the success rates for each classification scheme. The results are shown for 10-fold cross-validation and also for 50% of the total data being used for training (subsection 6.4.3). Superior performance of CP coupled with GD can be readily noted.

6.4.7 Performance for individual machines

Finally, we provide success rates for classification of ELM regions for individual machines. Results are given in Table 6.7 and were obtained using a 10-fold cross-validated inductive conformal predictor with the geodesic distance. Class-wise success rates are also given for each machine, where the two classes are H-mode region with small ELMs, denoted by 'S' and H-mode region with type I ELMs, denoted by 'Type I'.

For analyzing the results given in Table 6.7, the following characteristics of data need to be considered:

- Class imbalance: a two-class data set is considered imbalanced (or skewed) when one of the classes is heavily under-represented in comparison to the other class.

- Dispersion of data: degree to which the data points within a cluster are

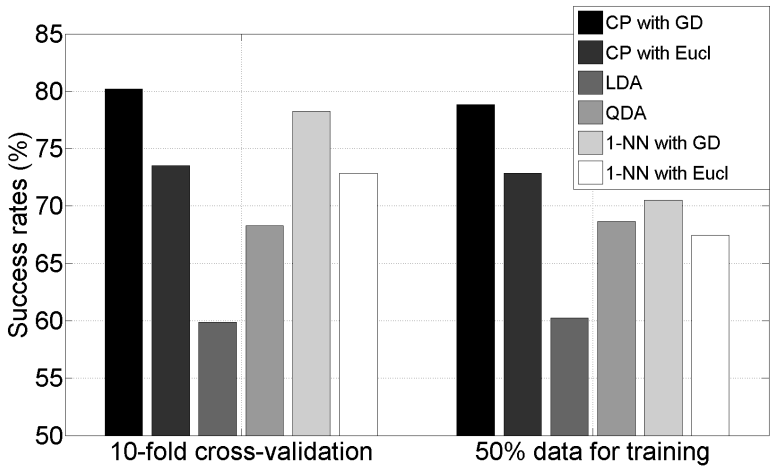


Figure 6.5: *Comparative SRs (%) for different classification schemes. The results are shown for both tenfold cross validation and random sampling of training data, where 50% of the total data are selected for training.*

dispersed over the feature space.

Class imbalance for each machine is listed in Table 6.8, while the statistics of dispersion within the class are given in Table 6.9. Dispersion is quantized by computing the mean distance to the nearest neighbor within the class. The larger the mean distance to the nearest neighbor, the larger is the spread within the class. Furthermore, higher is the standard deviation in the distance to the nearest neighbor for the samples, the lower is the likelihood of occurrence of localized clusters of a certain class. The distance to nearest neighbor is based on the GD between the probability distributions.

The data from JET are balanced and also have smallest within-class dispersion of all the machines. This can also be seen by visual inspection of Figure 6.2(d),

Machine	Class	SR(%) by class	SR (%)	ER (%)	AM (%)	CO (%)	CR (%)
JET	S	76.76	78.32	21.34	0.34	94.1	43.5
	Type I	79.88					
ASDEX	S	64.23	70.46	28.46	1.14	91.2	44.2
	Type I	76.7					
AUG	S	69.35	82.86	16.49	0.53	97.3	40.1
	Type I	96.37					
DIII-D	S	61.1	74.33	25.67	0	93.3	40.6
	Type I	87.55					

Table 6.7: *SR (%), ER (%), AM (%), CO (%) and CR (%) for ELM regime classification using ITPA data from JET, ASDEX, AUG and DIII-D, based on a 10-fold cross-validated ICP.*

Machine	Class	No. of samples per class	Ratio of class S to class Type I	Class balance
JET	S	800	45:55	Balanced
	Type I	980		
ASDEX	S	158	35:65	Partially unbalanced
	Type I	287		
AUG	S	85	15:85	Unbalanced
	Type I	498		
DIII-D	S	94	27:73	Partially unbalanced
	Type I	249		

Table 6.8: *Number of samples from each class for each machine, determining the respective class balance.*

Machine	Class	Mean distance to NN	Standard deviation
JET	S	1.75	2.33
	Type I	1.48	2.02
ASDEX	S	9.13	3.01
	Type I	9.83	2.93
AUG	S	12.78	2.79
	Type I	13.08	2.27
DIII-D	S	4.56	3.2
	Type I	2.69	2.0

Table 6.9: *Measure of dispersion within each class for each machine.*

where localized clusters of each class can be observed despite the limitations of the projected space. Hence classification performance for data from JET is high on the whole and also for each class individually. The data from ASDEX are not just partially unbalanced but also suffer from large within-class dispersion. This is also verified by the visual projection in Figure 6.2(c). As a consequence, classification performance is lower than that of other machines. The data from AUG are highly unbalanced in the favor of class type I, i.e. H-mode region with type I ELMs. As a result, the classification performance for class type I is very high and that of class S, i.e. H-mode region with small ELMs, is considerably lower. However, the high success rate for class type I outweighs the other effects, making the overall average success rate for this machine the highest. The success rate for class S for DIII-D is the lowest amongst all machines. Once again, this can be attributed to a partial imbalance of the class towards class type I and a higher dispersion within class S as compared to class type I.

6.4.8 Practical implications

The most significant contribution of this work is the development of an alternative methodology for classification of ELM types. It is noteworthy that the relative performance of the proposed method in contrast to other techniques is more important than the absolute value of the success rate. This is so because the absolute success rates are a stronger indicator of the quality of the chosen features and their discriminatory power for the problem at hand, rather than the goodness of the classification scheme. Incorporating additional features such as pedestal parameters, ELM frequency etc. is likely to significantly improve success rates and forms a part of the authors' ongoing work. Further, more accurate estimates of error bars and more precise information regarding the distribution of the uncertainties could be very useful for optimizing the classification performance. Suitability and advantages of the developed method have been demonstrated using the ITPA database, despite the limited information offered by it on the underlying probability distribution of the predictors. Having established the merits of the technique it can be rightfully argued that the method can be deployed as one of the layers in an embedded multi-layer classifier for ELM types.

6.5 Conclusions

We have presented an approach that offers a new perspective to the discrimination of ELM types, as an addition to the existing predominant phenomenological categorizations. The presented approach conducts pattern recognition using global plasma data while consistently taking into account uncertainties: first uncertainty of the data themselves and then of the classification results. This is important in order to increase the reliability of classifiers for resolving the underlying physics and for plasma control decisions.

The proposed distance-based conformal predictor classifier integrated with the geometric-probabilistic framework provides for an automated classifier for ELM types with high success rates and a figure of its own merit: confidence and credibility. Furthermore, it possesses dynamic learning capability and a mechanism for detection of ambiguities, which is advantageous over incorrect classification. The method is

generic and can be applied to other problems in nuclear fusion, such as disruption prediction. In addition, it is exportable to other application domains in signal and image processing. The method can also potentially help in quantifying the change in the behavior of ELMs in response to control and mitigation strategies.

Chapter 7

Classification of ELM types using distributions of global plasma parameters and inter-ELM time intervals

ELMs are a complex phenomena and hitherto, various types of ELMs have been identified and defined on an empirical and phenomenological basis (chapter 4). In this chapter, pattern recognition (chapter 2) is used for ELM classification with the following specific objectives:

- To provide a practical, standardized and automatic classification scheme for ELM types which can considerably reduce the effort of ELM experts in identifying ELM types
- To ensure that the ELM classification scheme is fast and has reasonably high accuracy
- To demonstrate that the distributions of ELM properties contain more information than the mean values alone

*The work presented here in chapter 7 has resulted in the following publications::

A.Shabbir, G.Hornung, G.Verdoolaege and JET contributors,"A classification scheme for edge-localized modes based on their probability distributions," *Review of Scientific Instruments*, vol.87, issue 11, no.11D404, 2016.

A.Shabbir, G.Hornung, J.-M.Noterdaeme, G.Verdoolaege and JET contributors,"Physical characterization of the JET operational space regarding ELMs by means of discriminant analysis," status: to be submitted.

	B_t (T)	I_p (MA)	n_e $(10^{19}m^{-2})$	P_{input} (MW)	δ_{avg}	Γ_{D_2} $(10^{22}s^{-1})$
Type I ELMy plasmas ($N = 74$)						
Range	1.4 - 3.0	1.4-3.0	3.2 - 9.9	7.6 - 22	0.21 - 0.50	0 - 8.9
μ	2.48	2.38	6.51	15.6	0.381	2.01
Median	2.67	2.48	6.38	15.6	0.416	1.17
σ	0.372	0.340	1.65	2.75	0.083	2.13
Type III ELMy plasmas ($N = 26$)						
Range	1.7 - 2.7	1.7-3.2	1.7 - 10.5	5.1 - 22	0.20 - 0.44	0 - 8.4
μ	2.22	2.32	6.57	12.0	0.387	4.13
Median	2.39	2.46	5.78	13.9	0.393	5.57
σ	0.288	0.417	2.41	5.58	0.0646	3.34

Table 7.1: *The overall ranges, class means, medians and standard deviations of the plasma parameters in the data set JET ELM-DB1.*

- To attempt to relate the ELM classification with the governing physical processes

- To explore more than one classification paradigm (parametric and non-parametric) and compare their usefulness for ELM classification and understanding of underlying physics.

7.1 Dataset: JET ELM-DBI

From the JET carbon wall plasmas, a selection of 69 type I, 26 type III and 5 type I high frequency (HF) ELMy discharges has been made to constitute the JET_CW ELMy database (JET ELM-DB1) reported in Appendix A.1: JET ELM-DBI. This is an extension of the data set used earlier by Webster *et al.* [159] for statistical characterization of ELM types. The analysis, in this work, has been restricted to time intervals in which the plasma conditions are quasi-stationary with approximately constant heating, gas fueling and central density. Further, all experiments dealing with ELM control and mitigation techniques have been excluded. The ranges, class means, medians and standard deviations of the global plasma parameters pertaining to the two classes i.e. type I ELMy plasmas and type III ELMy plasmas, are summarized in Table 7.1. The global plasma parameters considered herein are: vacuum toroidal field at $R = 2.96$ (B_t) (T), plasma current (I_p) (MA), line integrated edge density $(n_e)(10^{19}m^{-2})$, gas fueling$(\Gamma_{D_2})(10^{22}s^{-1})$,

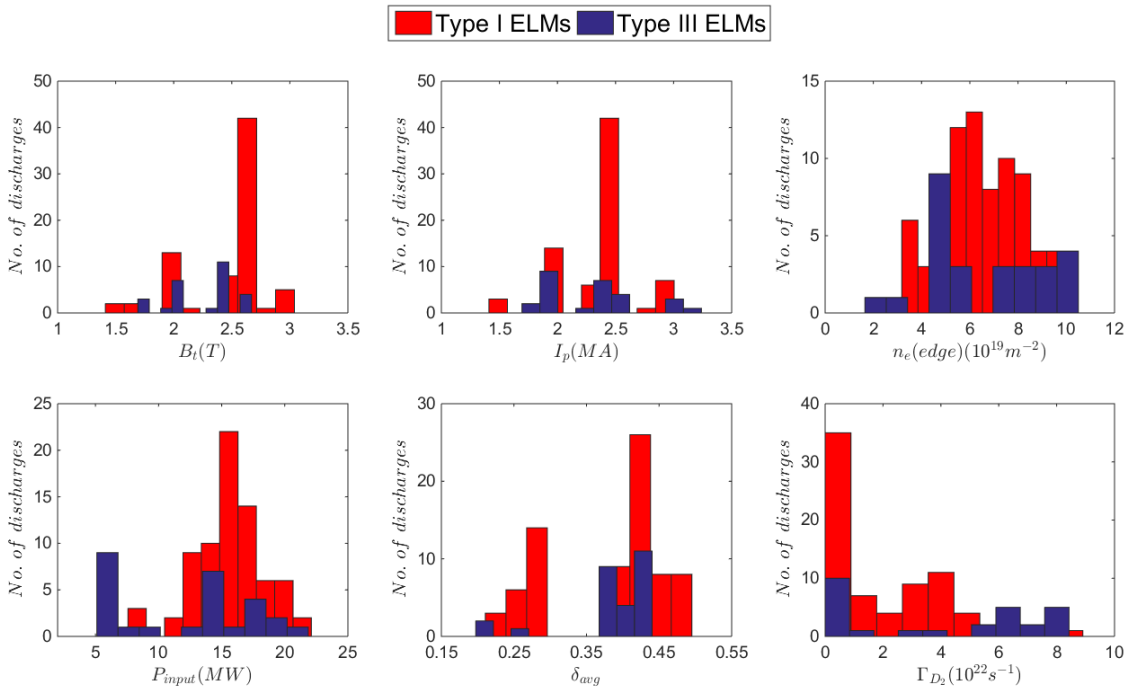


Figure 7.1: *Histograms of plasma parameters for the dataset JET ELM-DB1.*

input power $(P_{input})(MW)$ and average triangularity (δ_{avg}), where

$$P_{input} = P_{ohmic} + P_{NBI} + P_{ICRH}, \text{ and} \tag{7.1}$$

$$\delta_{avg} = \frac{\delta_{lower} + \delta_{upper}}{2}. \tag{7.2}$$

Histograms of plasma parameters for each class are presented in Figure 7.1. From a visual inspection of Figure 7.1, considerable overlap amongst the two classes can be readily observed.

7.2 Parametric classification

Discriminant analysis (DA) (see: Discriminant analysis (subsection 2.4.1)) is used for parametric classification. DA assumes that the data within each class is normally distributed and is applied here for the classification of type I and type III ELMs in the JET ELM-DBI dataset using global plasma parameters (B_t, I_p, n_e, P_{input}, δ_{avg} and Γ_{D_2}) as predictors. DA generates a decision surface in terms of the global plasma parameters which can be used for the classification of new discharges as well as the study of underlying physics.

7.2.1 Performance assessment

Typically, the performance of classification schemes is assessed by the *hold-out* procedure which involves splitting the data into two disjoint subsets called the *training* set and a *test* set. The training set is used for training the classifier and the test set is used for estimating the *success* rate (or alternatively the *error* rate) of the classifier. Though computationally simple, this can lead to possible over-fitting or under-fitting of training data leading to poor prediction performance (bias) on independent data sets. This is overcome by employing a leave-one-out cross validation method for performance estimation. For a dataset with N samples, N experiments are performed where $N-1$ samples are used for training and the remaining sample is used for testing. The leave-one-out cross validated success rate, which here is quoted as the percentage of samples that are correctly classified, is in effect, an estimation of the expected performance of the classifier on an unknown independent dataset. It is a yardstick for assessing the generalization capability of the classification scheme.

Further, the benchmark that we will use to characterize a classification model as useful is a 25% improvement over the rate of accuracy achievable by chance alone. The estimate of by chance accuracy is made by summing the squared percentage of samples belonging to each class.

$$By\ chance\ accuracy = P(I)^2 + P(III)^2, \tag{7.3}$$

where $P(I)$ is the prior probability of type I ELMs and $P(III)$ is the prior probability of type III ELMs.

Prior probability of type I ELMs	$74/10 = 0.74$
Prior probability of type III ELMs	$26/100 = 0.26$
By chance accuracy according to (7.3)	61.5%
25% increase over by chance accuracy	$Accuracy \geq 76.9\%$ or $Error < 23.1\%$

7.2.2 Classification performance

The discriminant analysis of the data set is performed in two stages. The first one is identifying the predictive capability of each plasma parameter and selecting the

	B_t (T)	I_p (MA)	n_e $(10^{19}m^{-2})$	P_{input} (MW)	δ_{avg}	Γ_{D_2} $(10^{22}s^{-1})$
Discriminant value	2.35	2.34	6.45	13.7	0.384	2.99
Resubstitution success (%)	73.0	74.0	74.0	82.0	74.0	82.0
Leave-one-out cross-validated success (%)	73.0	74.0	74.0	82.0	74.0	81.0

Table 7.2: *Predictive capability of single plasma paramters using linear discriminant analysis.*

parameters that are significantly relevant for ELM classification. The second one is deriving a specific discriminant function for demarcating the boundary between type I and type III ELMs in terms of global plasma parameters.

Linear discriminant analysis (LDA) (see: Discriminant analysis(subsection 2.4.1)) is performed on the discharges in JET ELM-DBI represented by each plasma parameter individually. Table 7.2 lists the resubstitution and leave-one-out cross validated success rates (%). Resubstitution success is the success rate (%) obtained on the training data and is in most cases an optimistic estimate. Leave-one-out cross validated success rate, as outlined earlier, makes the best use of the data for providing a realistic and a robust estimate of classification performance. For the results presented in Table 7.2 the estimated covariance matrices coincide with the variances of the two classes and the discriminant function is reduced to a discriminating value (DV). This DV, derived under the assumption of equal variances, is applied for classification. DV is given as:

$$DV = \frac{1}{2}(\mu_{class\ 1} + \mu_{class\ 2}). \tag{7.4}$$

Under the assumption of unequal variances, the analysis produces success rates similar to those presented in Table 7.2 (differences are $\sim 1\%$). Further, Table 7.2 reveals that the parameters P_{input} and Γ_{D_2} yield the highest success rates and hence may play the main role in the classification between the two classes.

Discriminant analysis is then performed on the linear and quadratic combinations of the plasma parameters, in order to further improve the success rate. The average and class-wise resubstitution success and leave-one-out cross validated success are given in Table 7.3. It can be noted that a linear combination of P_{input} and Γ_{D_2}

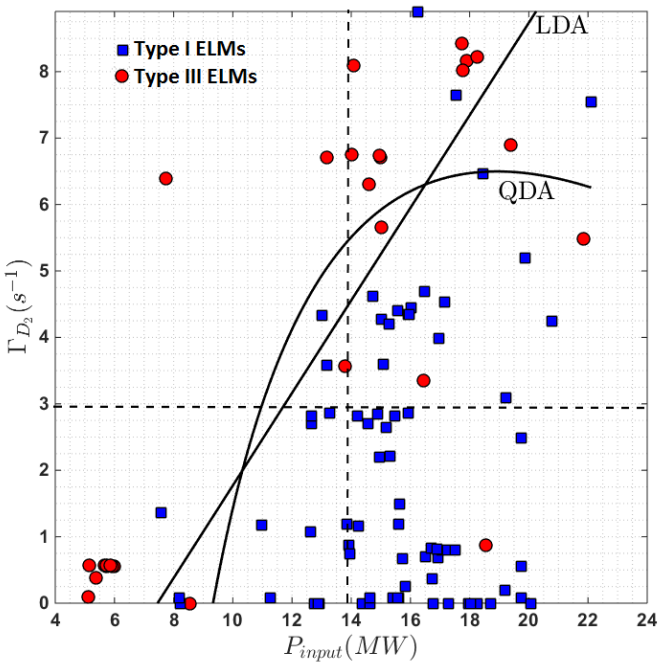


Figure 7.2: *The solid line and curve indicate the linear (LDA) and the quadratic (QDA) discriminant function respectively for type I and type III ELMs from JET ELM-DBI. Vertical and horizontal dashed lines mark the discriminating values for* P_{input} *and* Γ_{D_2}*, respectively.*

improves the average leave-one-out cross validated success rate to 91.0% from (81.0 - 82.0)% yielded by each of them individually. On the other hand, a quadratic combination of P_{input} and Γ_{D_2} increases the average success rate to 89.0%. This is further illustrated in Figure 7.2. It can be readily observed that the vertical and horizontal dashed lines discriminate the two classes poorly, whereas the solid lines, which are a function of P_{input} and Γ_{D_2}, better separate the two classes. A second important observation which can be made from the inspection of Figure 7.2 is that the hyperplane best separating the two classes has multiple solutions based on the minimization of classification error. Further, it can be noted that for 10.4 $MW \leq P_{input} \leq 16.5\ MW$ the difference between the quadratic and linear boundary is small ($\Delta\Gamma_{D_2} \leq 1.0 s^{-1}$). However, for $P_{input} > 16.5\ MW$, this difference is substantial.

Figure 7.3, presents the decrease in error rate (%) with the addition of other plasma parameters. An addition of the remaining 4 plasma parameters, B_t, I_p, n_e and δ_{avg} to P_{input} and Γ_{D_2} reduces the average error rate to 8% (alternatively, average success rate improves to 92%) for the linear combination of parameters and to 6% for the quadratic case. While the addition of Γ_{D_2} to P_{input} had reduced the error rate by a factor of ∼2, the addition of the remaining 4 parameters only

Plasma parameters		Resubstitution success (%)			Leave-one-out CV success (%)		
		I	III	Avg	I	III	Avg
P_{input}, Γ_{D_2}	LDA	93.2	84.6	91.0	94.6	80.8	91.0
	QDA	94.6	76.9	90.0	90.5	84.6	89.0
P_{input}, Γ_{D_2}, I_p	LDA	93.2	84.6	91.0	94.6	76.9	90.0
	QDA	94.6	80.8	91.0	91.9	80.8	89.0
P_{input}, Γ_{D_2}, δ_{avg}	LDA	93.2	84.6	91.0	94.6	80.8	91.0
	QDA	94.6	80.8	91.0	91.9	73.1	87.0
P_{input}, Γ_{D_2}, n_e	LDA	93.2	80.8	90.0	93.2	76.9	89.0
	QDA	94.6	73.1	89.0	90.5	80.8	88.0
P_{input}, Γ_{D_2}, B_t	LDA	93.2	73.1	88.0	90.5	80.8	88.0
	QDA	96.0	88.5	94.0	93.2	84.6	91.0
P_{input}, Γ_{D_2}, δ_{avg}, I_p	LDA	93.2	84.6	91.0	94.6	80.8	91.0
	QDA	94.6	88.5	93.0	96.0	80.8	92.0
P_{input}, Γ_{D_2}, δ_{avg}, n_e	LDA	91.9	88.5	91.0	91.9	30.8	14.0
	QDA	93.2	80.8	90.0	90.5	76.9	87.0
P_{input}, Γ_{D_2}, δ_{avg}, B_t	LDA	94.6	76.9	90.0	91.9	84.6	90.0
	QDA	97.3	96.2	97.0	97.3	73.1	91.0
$\boldsymbol{P_{input}}$, $\boldsymbol{\Gamma_{D_2}}$, $\boldsymbol{\delta_{avg}}$, $\boldsymbol{B_t}$, $\boldsymbol{I_p}$	**LDA**	97.3	84.6	94.0	**94.6**	**84.6**	**92.0**
	QDA	96.0	92.3	95.0	**94.6**	**88.5**	**93.0**
P_{input}, Γ_{D_2}, δ_{avg}, I_p, n_e	LDA	93.2	84.6	91.0	93.2	80.8	90.0
	QDA	93.2	88.5	92.0	91.9	80.8	89.0
P_{input}, Γ_{D_2}, δ_{avg}, B_t, n_e	LDA	94.6	80.8	91.0	90.5	80.8	88.0
	QDA	97.3	0.00	98.0	97.3	76.9	92.0
$\boldsymbol{P_{input}}$, $\boldsymbol{\Gamma_{D_2}}$, $\boldsymbol{n_e}$, $\boldsymbol{B_t}$, $\boldsymbol{I_p}$	**LDA**	97.3	84.6	94.0	**94.6**	**84.6**	**92.0**
	QDA	97.3	88.5	95.0	**94.6**	**92.3**	**94.0**
P_{input}, Γ_{D_2}, δ_{avg}, B_t, I_p, n_e	LDA	96.0	92.3	95.0	94.6	84.6	92.0
	QDA	96.0	96.2	96.0	96.0	88.5	94.0

Table 7.3: *Average and class-wise resubstitution success (%) and leave-one-out cross validated success (%) for a linear and quadratic combination of plasma parameters obtained by LDA and QDA, respectively.*

lowers it further by 1% for LDA and 5% for QDA. It is noteworthy that the error rate for type III ELMs reduces by $\sim$ 4% for both LDA and QDA whereas the error rate for type I ELMs remains unchanged for LDA and lowers by $\sim$ 5% for QDA. However, this reduction in error rates comes at the expense of an increased model complexity. An increase in the number of parameters in the discriminant function increases model complexity which in turn leads to higher variance (bias-variance tradeoff) as well as a less tractable model.

It can be noted from Table 7.3 and Figure 7.3 that from the various models analyzed, the linear combination of P_{input}, Γ_{D_2}, B_t, I_p along with either n_e or δ_{avg}, can be considered as the best models as they yield high average and class wise success

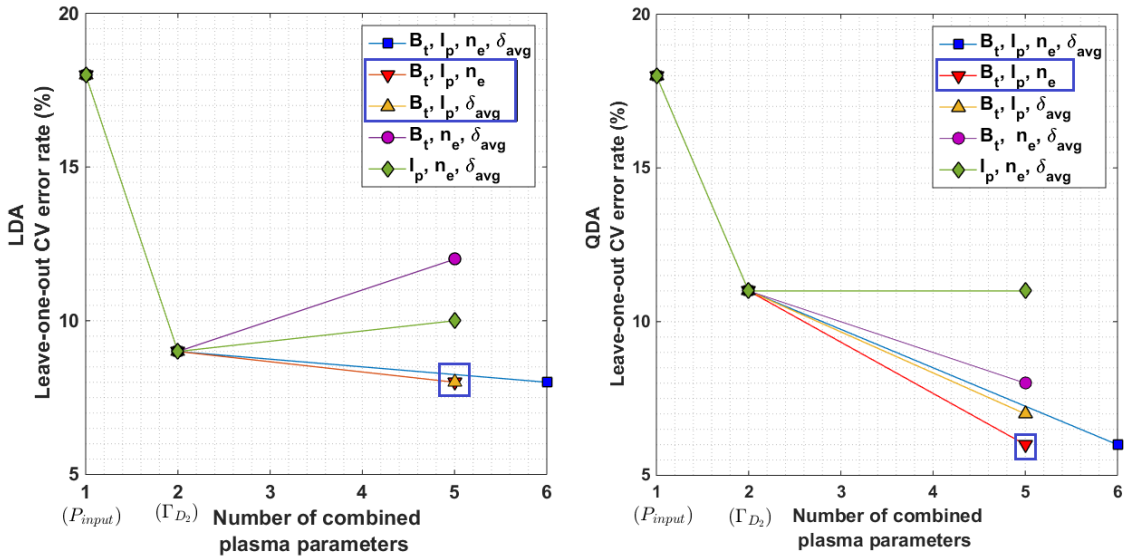


Figure 7.3: *Leave-one-out cross validated error rate versus the number of combined plasma parameters using (a). LDA, (b). QDA*

rates with the least number of parameters. These two models are highlighted in Table 7.3 and Figure 7.3. The quadratic combination of P_{input}, Γ_{D_2}, B_t, I_p and n_e gives the highest success rate amongst all analyzed models. However, the quadratic model is significantly more complex, less intuitive and less tractable, than the linear counterpart. However, if the primary goal is correct classification of a new discharge, then this quadratic model can be slightly advantageous compared to the linear ones.

7.2.3 Separation hyperplane for type I and type III ELMs

The mathematical form for the linear discriminant functions derived for the classification of type I and III ELMs is presented in Table 7.4. The classification success rates for these linear separating hyperplanes (boundary) are provided in Table 7.3. For each of the three discriminant functions, L1, L2 and L3 given in Table 7.4, if the left hand side of the expression is less than the constant on the

	Linear discriminant functions	Wilks' Λ
L1	$P_{input} - 1.41\Gamma_{D_2} = 7.47$	0.60
L2	$P_{input} - 1.25\Gamma_{D_2} + 7.06B_t - 8.81I_p + 0.70n_e = 8.75$	0.53
L3	$P_{input} - 0.765\Gamma_{D_2} + 12.4B_t - 10.7I_p - 26.1\delta_{avg} = 3.96$	0.47

Table 7.4: *Linear separation hyperplanes (boundary) for type I/III ELMs, in terms of global plasma parameters. The corresponding classification success rates (%) are provided in Table 7.3. Wilks' Λ indicates the goodness of fit of each discriminant function.*

.

Predictors	L1	L2	L3
P_{input}	0.53	0.47	0.41
Γ_{D_2}	-0.46	0.38	-0.36
B_t	-	0.33	0.30
I_p	-	0.078	0.070
n_e	-	-0.014	-
δ_{avg}	-	-	-0.030

Table 7.5: *Correlation between plasma parameters (predictors) and the discriminant functions L1, L2 and L3.*

right hand side, type III ELMs are expected (or obtained). Vice versa, if the left hand side of the discriminant function is more than the constant on the right hand side, type I ELMs are expected.

Assessing model fit

The goodness of fit of the estimated discriminant functions to the hyperplane separating type I and III ELMs is assessed using a statistic called Wilks' lambda (Λ) [160]. The Wilks' Λ estimates for L1, L2 and L3 are given in Table 7.4. Theoretically, the closer Wilk's Λ is to 0, the better is the model fit. However, for practical purposes a value of Wilk's Λ less than or equal to 0.63 implies a reasonably good fit.

Assessing the contribution of individual parameters

Table 7.5 presents the estimates of the correlations between each plasma parameter in the model and the discriminant functions. These estimates allow us to see how closely a parameter is related to each discriminant function and provide an indication of the importance of each parameters unique contribution to the discriminant function. The results in Table 7.5 suggest that P_{input}, Γ_{D_2} and B_t contribute the most to the hyperplanes separating type I and III ELMs.

Illustration of the use of discriminant function

Given L2:

$$P_{input} - 1.25\Gamma_{D_2} + 7.06B_t - 8.81I_p + 0.70n_e = 8.75$$

If $B_t = 2T$, $I_p = 2MA$, $n_e = 6.0 \times 10^{19}m^{-2}$ and $P_{input} = 15MW$, a gas fuelling rate of $5.56 \times 10^{22}s^{-1}$ and above will result in type III ELMs. However, if P_{input} is

increased to 20 MW, a a gas fuelling rate of $9.56 \times 10^{22} s^{-1}$ and above will lead to type III ELMs. This behavior is well-aligned with known physics.

7.3 Non-parametric classification

Parametric classification using discriminant analysis offers various advantages. However, its potential is challenged by two inherent assumptions: data within classes is assumed to be normally distributed and in case of LDA, the classes are considered to be sharing a common covariance matrix. Furthermore, it has been shown that working with distributions of plasma parameters and ELM properties has significant benefits over working with mean values alone. Discriminant analysis operates on the mean values of the parameters and does not accommodate for the uncertainty on the parameters.

Hence, parametric classification using k-nearest neighbors (kNN) classifier is performed next on JET ELM-DBI dataset. In the first stage, kNN classifies ELMs using the distributions of global plasma parameters which allows for a direct comparison with the results obtained with DA. In the second stage, non-parametric classification is performed on the basis of the distributions of inter-ELM time intervals (Δt_{ELM}), also interchangeably referred to as the *waiting times*. It will be shown that this provides an appreciable improvement over the classification of ELMs using the average ELM frequency (f_{ELM}).

7.3.1 Using global plasma parameters

Table 7.6 presents the leave-one-out cross-validated success rates (%) for 1-nearest neighbour (1-NN) classification of type I and type III ELMs from JET ELM-DBI using the distributions of the aforementioned global plasma parameters, i.e. B_t, I_p, n_e, P_{input}, δ_{avg} and Γ_{D_2}, as predictors. As in chapter 5, we assume that the error bars associated with each plasma parameter pertain to a statistical uncertainty in the data, specifically that it represents a single standard deviation. According to the principle of maximum entropy the underlying probability distribution is Gaussian with the measurement and its error bar constituting the mean (μ) and the standard deviation (σ), respectively [161].

It can be noted from Table 7.6 that for each set of predictors, the GD-based

Plasma parameters		Distance measure	Leave-one-out CV success (%)		
			I	III	Avg
P_{input}, Γ_{D_2}	μ	Eucl.	85.1	46.2	75.0
	μ, σ	GD	86.5	73.1	83.0
P_{input}, Γ_{D_2}, B_t I_p, n_e	μ	Eucl	89.2	69.2	84.0
	μ, σ	GD	93.2	80.8	90.0
P_{input}, Γ_{D_2}, B_t I_p, δ_{avg}	μ	Eucl	89.2	69.2	84.0
	μ, σ	GD	95.9	84.6	93.0
P_{input}, Γ_{D_2}, B_t I_p, n_e, δ_{avg}	μ	Eucl	89.2	69.2	84.0
	μ, σ	GD	95.9	84.6	93.0

Table 7.6: *Leave-one-out cross-validated (CV) classification success rates (%) for the classification of type I and type III ELMs from JET ELM-DBI using global plasma parameters as predictors and 1-nearest neighbour (1-NN) classifier. Euclidean distance based 1-NN is used for classifying on the basis of the mean (μ) values of plasma parameters and both Euclidean distance based 1-NN and GD-based 1NN are used for classifying on the basis of distributions (μ, σ) of plasma parameters.*

classification using the distributions of the plasma parameters performs significantly better than the classification based on the mean values of the parameters. A comparison with the success rates provided in Table 7.3 reveals that both DA and kNN classify with reasonably high accuracy and their classification performance is comparable. This renders the choice of a classifier application dependent. kNN as a completely non-parametric approach makes no assumptions about the shape of the decision boundary and is expected to outperform LDA when the decision boundary is highly non-linear. On the other hand, kNN does not provide an interpretable model.

7.3.2 Using inter-ELM time intervals

7.3.2.1 Extraction of ELM temporal characteristics

JET plasmas typically contain more than a hundred ELMs in the stationary phase of the discharge. Most ELMs are not indexed and thus a thorough statistical analysis of their properties is far from easy. A robust ELM detection algorithm is developed for the extraction of inter-ELM time intervals. The algorithm can also be used for estimating other temporal characteristics such as ELM durations, ELM crash time as well as the synchronization of ELMs for ELM energy loss estimation (Chapter 8). ELM detection algorithms typically examine the radiation associated with ELMs, using a threshold or a combination of thresholds to signal the start and end of an

Type of discharges	ELM signature used for analysis
JET CW	D_α emissions at the inner divertor (p/scm^2sr)
JET ILW	$Be\ II\ 527nm$ photon flux at the inner divertor (p/scm^2sr)
AUG (full W wall)	Scrape-off layer current measured via shunt resistances at the outer divertor (kA)

Table 7.7: *Type of discharge and the corresponding ELM signature signal used for analysis and as input to the ELM detection algorithm.*

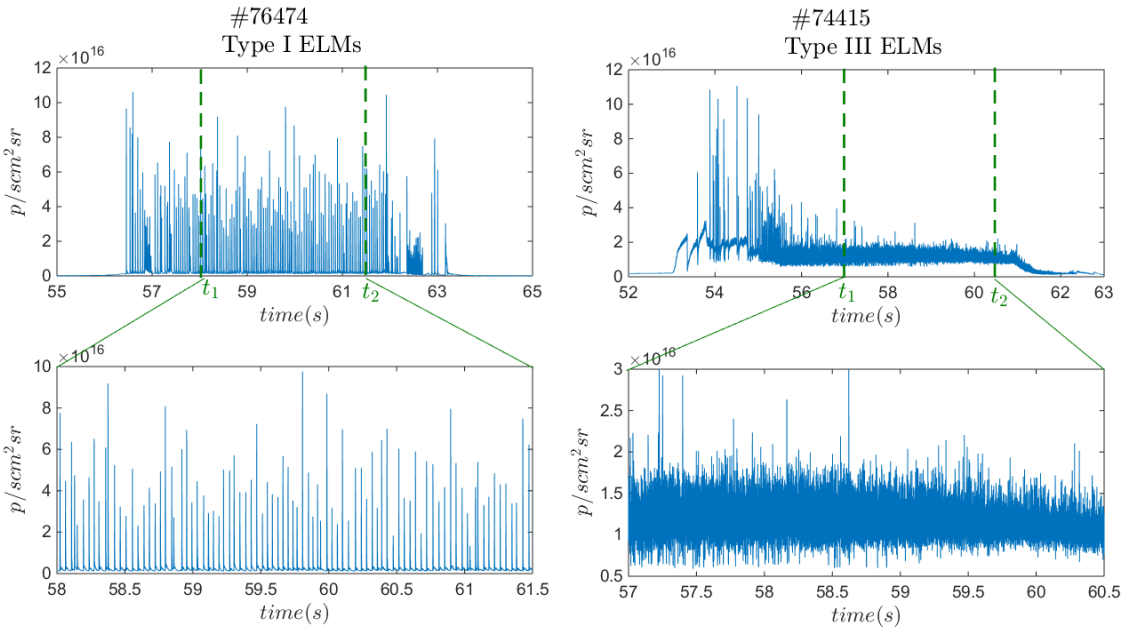


Figure 7.4: *D_{alpha} signal (ELM signature) for a type I and a type III ELMy discharge from JET ELM-DBI. The analysis time interval delimited by t_1 and t_2 is marked by dashed lines.*

ELM [159][162]. In those respects, our detection algorithm is the same and performs a series of sequential operations on the given input.

STEP 1: The algorithm requires as input, an ELM signature signal and a pre-set analysis time interval which pertains to quasi stationary plasma conditions. Corresponding to the three categories of discharges, i.e. JET-CW plasmas, JET-ILW plasmas and AUG (full W wall) plasmas, Table 7.7 lists the ELM signature signals which are used as input to the ELM detection algorithm. For each discharge in the database, the time at which the time series analysis starts (t_1) and ends (t_2) is manually affixed. As an illustration, the D_α signal for two discharges from JET ELM-DBI is presented in Figure 7.4. The analysis time interval $[t_1, t_2]$ is indicated by dashed lines.

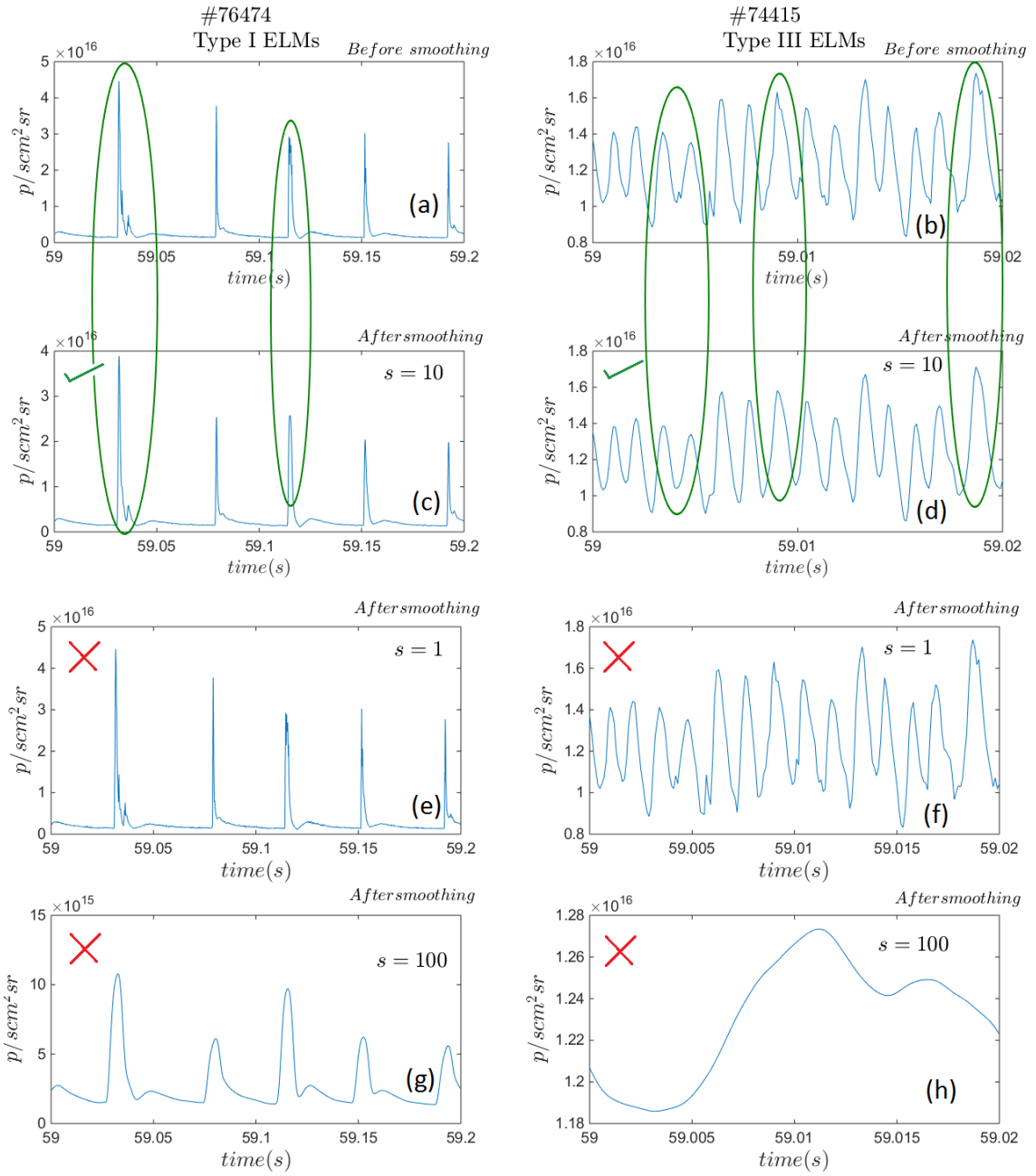


Figure 7.5: *Application of LOWESS smoothing to the ELM signature signal. (a)-(b). D_α signal for discharge #76474 (Type I ELMs) and #74415 (Type III ELMs). (c)-(d). D_α signal after smoothing with $s = 10$. (e)-(f). D_α signal after smoothing with $s = 1$. (g)-(h). D_α signal after smoothing with $s = 100$.*

STEP 2: The ELM signature signal is smoothed using *locally weighted least squares smoothing* (LOWESS) [163]. LOWESS smooths the ELM signature time series via local regression using weighted linear least squares fitting to the data points in the smoothing window. LOWESS smoothing, applied to a type I ELMy discharge and a type III ELMy discharge from JET ELM-DB1, is illustrated in Figure 7.5 (a)-(d). The width of the smoothing window (s)

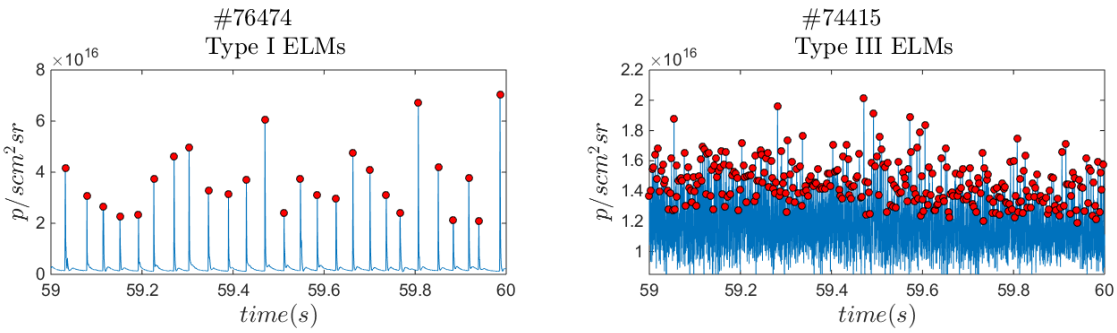


Figure 7.6: *ELM identification via peak detection.*

(also called *span*) specifies the number of closest neighbors of each data point that constitute the moving smoothing window. By trial and error, s has been set to 10 (Figure 7.5(c)-(d)) such that the smoothing stage facilitates subsequent processing, improves signal to noise ratio and reduces high frequency inter-ELM fluctuations. Figure 7.5 (e)-(h) illustrates the effect of choosing a large and a very small value of s. It can be seen that choosing $s = 1$ is ineffective as the smoothing effect is almost negligible. In contrast, choosing $s = 100$ leads to a loss of information as it causes an over-smoothing.

STEP 3: ELMs are identified by detecting peaks in the ELM signature signal. Peaks are detected by using the *zero-derivative* method as the first derivative of a peak has a downward going zero-crossing at the peak maximum. ELM identification is illustrated in Figure 7.6.

STEP 4: ELM start ($t_{ELM_{start}}$) and stop times ($t_{ELM_{end}}$) are determined. ELM is considered to end when the the amplitude of the ELM signature signal falls to ∼20-40% of the ELM peak value. The corresponding time is noted as $t_{ELM_{end}}$. Similarly, the time instance before the ELM peak time at which the amplitude of the ELM signature signal is ∼20-40% of the peak amplitude, is noted as $t_{ELM_{start}}$.

OUTPUT: For each ELM i, in a discharge with N ELMs, $t_{ELM_{start}}$, $t_{ELM_{peak}}$ and $t_{ELM_{end}}$ are obtained. The waiting time (Δt_{ELM}) is computed as:

$$\Delta t_{ELM} = t_{ELM_{end_{i+1}}} - t_{ELM_{start_i}}. \tag{7.5}$$

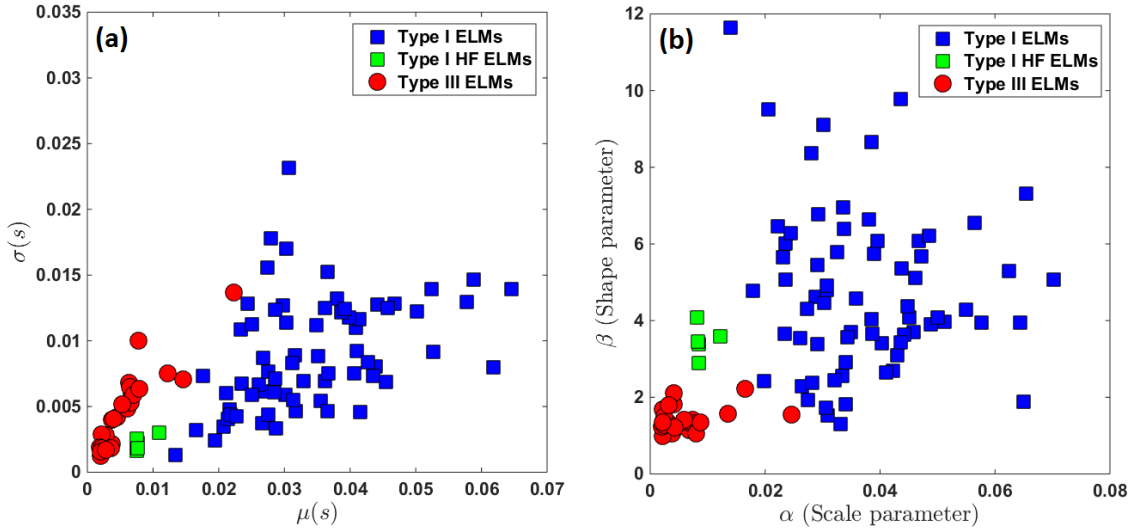


Figure 7.7: *Maximum-likelihood parameter estimates for (a). Gaussian distribution fit, (b). 2-parameter (2P) Weibull distribution fit to the ELM waiting times (Δt_{ELMs}) from JET ELM-DBI.*

7.3.2.2 Maximum-likelihood parameter estimation

Gaussian and 2-parameter (2P) Weibull distributions are used for modeling the N waiting times extracted from each discharge. Webster *et al.* [159] has recently shown that based on experimentally motivated assumptions, the 3-parameter (3P) Weibull distribution is a good model for capturing the waiting time statistics. However, the GD between 3P Weibull distributions does not have an analytical solution. Hence, for ensuring that the developed classification system is computationally efficient and as a first approximation, the 2P Weibull distribution is used herein.

Both Gaussian and 2P Weibull distributions have free parameters that require estimation. To this end, we consider the likelihood function for the probability of the data given the model being considered (M) and parameters (θ),with

$$L(\theta) = P(\{\Delta t_{ELM}\}|M, \theta), \tag{7.6}$$

where, $P(\{\Delta t_{ELM}\}|M, \theta)$ is the probability of observing the set of waiting times $\{\Delta t_{ELM_N}\}$, given the assumption of a distribution M (where M is Gaussian or 2P Weibull), with parameters θ. The parameters that maximize $L(\theta)$ are the maximum likelihood (ML) estimates of the distribution parameters.

ML estimates of the parameters of a Gaussian distribution (μ, σ) and the parameters of a 2P Weibull distribution (β, α) fit to the waiting times extracted

from the discharges in JET ELM-DBI are shown in Figure 7.7.

An examination of Figure 7.7, provides various insights. Figure 7.7(a) suggests that there is a positive linear correlation between mean and the standard deviation of the waiting times. This implies that type I ELMs, which typically have a higher mean waiting time, tend to have a wider distribution (i.e higher standard deviation) than type III ELMs. Furthermore, both the mean waiting time and its standard deviation appear to be discriminators of ELM type, especially for the discharges which lie at the boundary between type I and type III ELMs. For example, type I HF ELMs have mean waiting times which are smaller than typical type I ELMs but are more similar to type III ELMs. However they tend to have a smaller standard deviation than the standard deviation of type III ELMs with similar mean waiting times.

Figure 7.7(b) indicates that β and α are also both discriminators for ELM type. Type I ELMs typically have a higher value for α than type III ELMs. Also, the information in β appears useful for correctly classifying type I HF ELMs, since they have a higher value of β than the type III ELMs with similar values of α.

7.3.2.3 Classification performance

Table 7.8 presents the classification success rates (%) for k-nearest neighbor classification of ELM types from JET ELM-DBI using the distribution of ELM waiting times as a predictor. The following key observations can be made on the results outlined in Table 7.8:

- ELM classification using GD-based kNN with the Gaussian distribution parameters (μ, σ) as predictors (k=1), and the 2P Weibull distribution parameters (β, α) as predictors (k=3), yields the highest class wise and average success rates.

- The improvement in success rates when the distribution (μ, σ) of the waiting times is used as a predictor, over the success rate obtained with only the mean waiting time, confirms that the complete distribution of the waiting times contain more information than the mean values alone.

- With Gaussian distribution parameters (μ, σ) as predictors, GD-based kNN yields higher success rates than Euclidean distance based kNN. This reconfirms

Predictors	Distance measure	k	Leave-one-out CV success (%)		
			I	III	Avg
μ	Eucl	1	95.9	84.6	93.0
(μ, σ)	Eucl	1	95.9	84.6	93.0
$(\boldsymbol{\mu}, \boldsymbol{\sigma})$	**GD**	**1**	**97.3**	**96.2**	**97.0**
(β, α)	Eucl	1	94.6	80.8	91.0
(β, α)	GD	1	97.3	92.3	96.0
μ	Eucl	3	95.9	88.5	94.0
(μ, σ)	Eucl	3	95.9	88.5	94.0
(μ, σ)	GD	3	94.6	96.2	95.0
(β, α)	Eucl	3	94.6	92.3	94.0
$(\boldsymbol{\beta}, \boldsymbol{\alpha})$	**GD**	**3**	**97.3**	**96.2**	**97.0**
μ	Eucl	5	91.9	84.6	90.0
(μ, σ)	Eucl	5	91.9	84.6	90.0
(μ, σ)	GD	5	95.9	92.3	95.0
(β, α)	Eucl	5	87.8	88.5	88.0
(β, α)	GD	5	94.6	92.3	94.0

Table 7.8: *Class wise and average classification success rates (%) for type I and III ELMs from JET ELM-DBI using mean value and distributions of ELM waiting times as predictors together with a kNN classifier.*

that GD is a natural and a well suited similarity measure for comparing probability distributions. Likewise, with 2P Weibull distribution parameters (β, α) as predictors, GD exhibits a considerably superior performance over the Euclidean distance.

In order to further elucidate the capacity of the classification scheme, the region in the Gaussian parameter space overlapped by type I and type III ELMs is identified and indicated in Figure 7.8 with a dashed rectangle. The overlapping region spans an ELM frequency (f_{ELM}) range of $60 - 135\ Hz$. It can be seen from Table 7.9 that the GD-based 1-nearest neighbour (1-NN) classifier using the distribution of waiting times correctly classifies all 11 plasmas lying in the overlapping region. Whereas, the 1-NN classifier using only mean waiting times as predictors, incorrectly classifies 3 of the plasmas hailing from the overlap region. This helps to illustrate that complete distributions of ELM properties encompass more information which, if effectively exploited, can result in an improvement in the system performance.

The distribution of waiting times has rendered itself as a very important sorting quantity for ELM types. However, a further addition of the distribution of other ELM properties (such as ELM energy loss (W_{ELM}) or ELM durations (τ_{ELM})) to

the set of predictors can possibly further improve the performance.

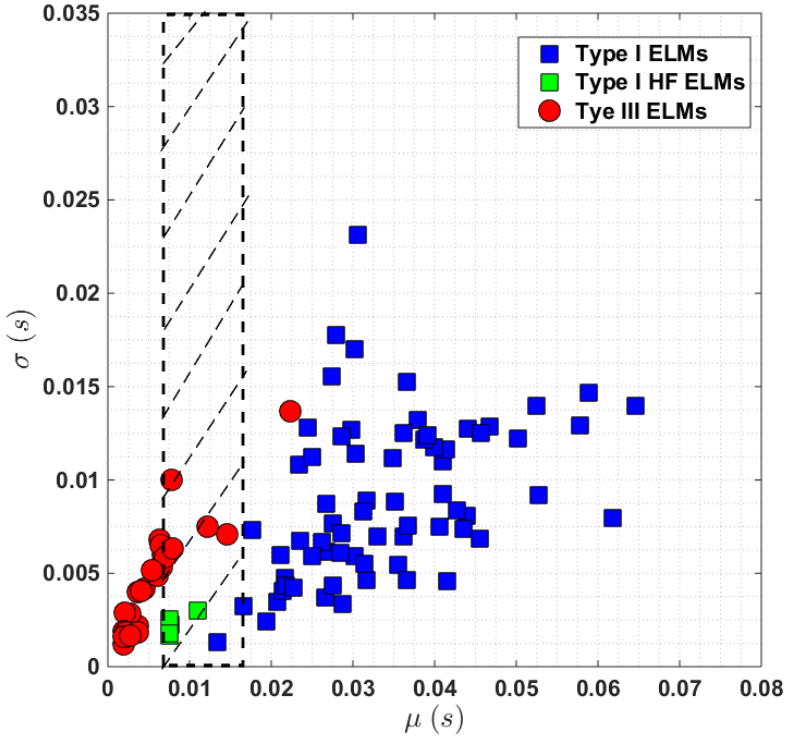


Figure 7.8: *Maximum-likelihood parameter estimates for a Gaussian distribution fit to the ELM waiting times (Δt_{ELMs}) from JET ELM-DBI. The region of the parameter space overlapped by type I and type III ELMs is indicated by a dashed rectangle.*

Shot number	ELM type	Predictors and distance measure	
		μ (Eucl)	(μ, σ) (GD)
56740	type I	✓	✓
67761	type I	×	✓
66109	type I HF	✓	✓
66108	type I HF	✓	✓
66107	type I HF	✓	✓
66106	type I HF	×	✓
66105	type I HF	✓	✓
70136	type III	✓	✓
74661	type III	✓	✓
74519	type III	✓	✓
74429	type III	×	✓

Table 7.9: *Class label (ELM type) predicted by GD-based 1-NN and Euclidean distance based 1-NN classifier for the discharges belonging to the overlap region in the Gaussian parameter space, indicated in Figure 7.8. The second column specifies the actual ELM type and the third and fourth column indicate if the ELM type is correctly (check mark) or incorrectly (cross mark) predicted.*

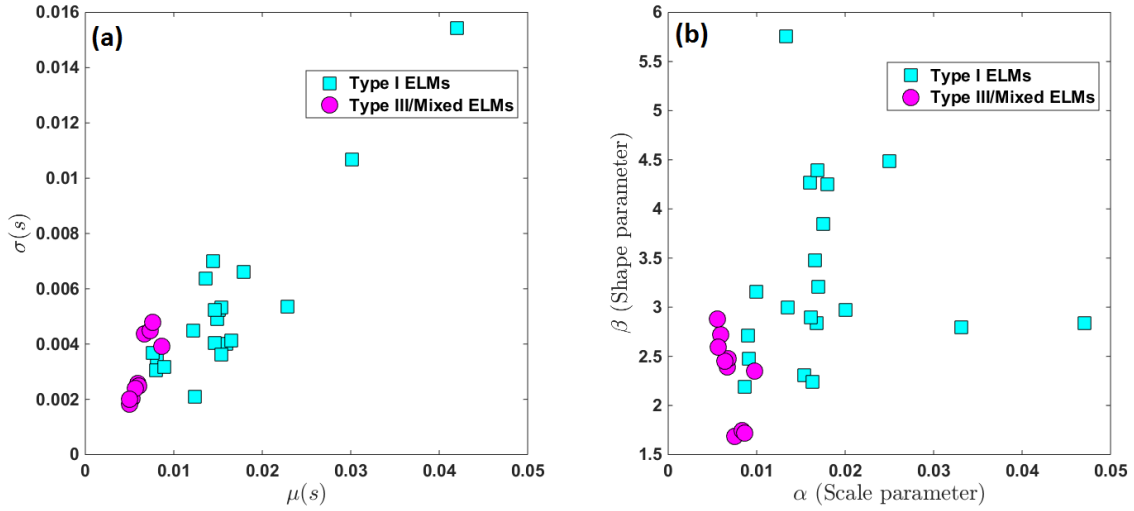


Figure 7.9: *Maximum-likelihood parameter estimates for (a). Gaussian, (b). 2P-Weibull distribution fit to the ELM waiting times (Δt_{ELMs}) from AUG ELM-DBI.*

7.3.2.4 AUG ELM-DBI: Classification performance

From the full-W wall experiments at AUG, a selection of 20 type I and 10 type III/mixed ELMy discharges has been made to constitute the AUG_CW ELMy database (AUG ELM-DB1) reported in Appendix A.1: AUG ELM-DBI.

The analysis has been restricted to time intervals in which the plasma conditions are quasi-stationary with approximately constant heating, gas fueling and central density. Further, all experiments dealing with ELM control and mitigation techniques have been excluded. For each discharge, inter-ELM time intervals are extracted using the ELM detection algorithm presented in the previous section. Maximum-likelihood estimates of Gaussian and 2P-Weibull distribution fits to the waiting times from AUG ELM-DBI are presented in Figure 7.9. As in the previous section, an examination of Figure 7.9 (a) reveals that both the mean (μ) and the standard deviation (σ) of the Gaussian distribution fit appear to contribute to the discrimination between type I and type III/small ELMs.

Table 7.10 presents the classification success rates (%) for kNN classification of ELM types from AUG ELM-DBI. In consistence with the results obtained with JET ELM-DBI, GD based classification using complete distribution of the waiting times yields the highest classification accuracy.

Predictors	Distance measure	Leave-one-out CV success (%)		
		I	III	Avg
μ	Eucl	90.0	80.0	86.7
(μ, σ)	Eucl	90.0	80.0	86.7
$\boldsymbol{(\mu, \sigma)}$	**GD**	**100**	**90.0**	**96.7**
(β, α)	Eucl	90.0	80.0	86.7
$\boldsymbol{(\beta, \alpha)}$	**GD**	**100**	**90.0**	**96.7**

Table 7.10: *Class wise and average classification success rates (%) for type I and III ELMs from AUG ELM-DBI using mean value and distributions of ELM waiting times as predictors and 1-NN classifier.*

7.4 Conclusions

In this chapter, a DA based parametric classification scheme and a kNN based non-parametric classification scheme have been presented and applied for the classification of ELMs from a database of JET plasmas and a small database of AUG plasmas. Each classification paradigm offers unique advantages and indeed the choice of a suitable ELM classifier is application dependent.

While DA requires more theoretical work before it can be applied on the probabilistic manifolds, classification using GD-based kNN clearly show that the complete distributions of global plasma parameters and ELM waiting times contain more useful information than the average parameter value. The presented schemes are fast and have been demonstrated to be capable of correctly predicting the ELM behavior even in those cases where the ELM type is not immediately clear.

CP classifier which offers certain well defined benefits has demonstrated its potential as a non-parametric classifier in chapter 6, where it has been applied for the classification of type I and small ELM from an international database. Herein, as ELM properties have been extracted from an assembled database from JET and AUG, non-parametric classification is performed by the simpler kNN classifier. Indeed, the CP classifier can be applied next for reaping the advantages that it offers.

Chapter 8

Correlation analysis for energy losses, waiting times and durations of type I edge-localized modes in JET*

A. Shabbir[1,2], G.Verdoolaege[1,3], G.Hornung[1],O.J.W.F. Kardaun[2],H.Zohm[2]

and JET contributors[1]

EUROfusion Consortium, JET, Culham Science Centre, Abingdon, OX14 3DB, UK

[1]Department of Applied Physics, Ghent University, Ghent 9000, Belgium

[2]Max-Planck Institute for Plasma Physics, Garching 85748, Germany

[3]Laboratory for Plasma Physics – Royal Military Academy (LPP – ERM/KMS), Brussels 1000 , Belgium

Abstract Several important ELM control techniques are in large part motivated by the empirically observed inverse relationship between average ELM energy loss and ELM frequency in a plasma. However, to ensure a reliable effect on the energy released by the ELMs, it is important that this relation is verified for individual ELM events. Therefore, in this work the relation between ELM energy loss (W_{ELM}) and waiting time (Δt_{ELM}) is investigated for individual ELMs in a set of ITER-like wall plasmas in JET. A comparison is made with the results from a set of carbon-wall and nitrogen-seeded ITER-like wall JET plasmas. It is found that the correlation

[1]See the Appendix of F. Romanelli et al., Proceedings of the 25th IAEA Fusion Energy Conference 2014, Saint Petersburg, Russia.

*The work presented here in chapter 8 is a preprint of a manuscript currently under review.

between W_{ELM} and Δt_{ELM} for individual ELMs varies from strongly positive to zero. Furthermore, most of the unseeded JET ILW plasmas have ELMs that are followed by a second collapse phase referred to as the slow transport event (STE). The effect of the STEs on the distribution of ELM durations is studied, as well as their influence on the correlation between W_{ELM} and Δt_{ELM}. A high correlation between W_{ELM} and Δt_{ELM}, comparable to CW plasmas is only found in nitrogen-seeded ILW plasmas. Finally, a regression analysis is performed using plasma engineering parameters as predictors for determining the region of the plasma operational space with a high correlation between W_{ELM} and Δt_{ELM}.

8.1 Introduction

Standard high confinement (H-mode) regimes in tokamaks are characterized by the existence of an edge transport barrier (ETB) in a narrow edge region inside the separatrix. Steep pressure gradients in the ETB lead to magnetohydrodynamic (MHD) instabilities called the *edge-localized modes* (ELMs) [164][3]. ELMs are intense, short duration, repetitive events that cause a partial collapse of the ETB and result in sudden expulsion of energy and particles from the plasma edge. On the one hand, ELMs pose a serious concern as they can cause high transient heat loads on the plasma-facing components (PFCs). On the other hand, they are crucial for regulating the core concentration of impurities, in particular, tungsten (W) which is produced by plasma-wall interactions at the divertor target.

Given the importance of ELMs for the successful operation of next-step fusion devices, a large array of ELM control and mitigation techniques have emerged [119][165]. Typically, ELM losses are influenced either by a complete suppression of the ELMs in regimes where an alternate mechanism replaces the energy and particle transport, or by increasing the ELM frequency (f_{ELM}) over its natural value (*ELM pacing*), so that the ELM losses become smaller. The effectiveness of the latter method in reducing the peak ELM energy flux (q_{max}) at the ITER divertor may be dampened in the wake of the experimentally observed linear dependence of the effective ELM energy deposition area (A_{ELM}) on ELM size (W_{ELM}) [140][166][141].

However, Loarte *et al.* [142] notes, that while the broadening of A_{ELM} certainly expands the operational regime of uncontrolled ELMs, for conditions in which the

uncontrolled ELMs would exceed the limits posed by divertor erosion, ELM control will be necessary at ITER. Secondly, the processes that lead to the broadening of A_{ELM} at the divertor will also have a similar effect on the scrape-off layer (SOL). This will inevitably result in an increase in the energy deposited on ITER's main wall which will consist of Beryllium (Be) PFCs. Be in contrast to the divertor material W, has a much lower erosion threshold which makes it highly likely that for some conditions the erosion limit of the first wall could constrain uncontrolled ELM operation.

Further, the recent ELM pacing experiments at DIII-D using lithium granules in contrast to frozen deuterium pellets, report on a reduction of the q_{max} at the outer strike point [144]. This result not only suggests the possibility of reducing q_{max} at ITER by non-fuel pellet injection but also presents an added advantage of de-coupling ELM pacing from plasma fueling.

Furthermore, in addition to the protection of PFCs, ELM control requirements at ITER have been recently revised to include W impurity control [135][142]. Excessive W concentration in the core can lead to severe central radiation losses which can affect the H-mode performance and in extreme cases result in a radiative collapse [167]. Experimental observation at JET [145] and AUG [168] have shown that a sufficiently high f_{ELM} will be required in ITER for maintaining an appropriate W concentration in the plasma.

ELM pacing [169][170], a leading candidate for controlling (W_{ELM}) in ITER, relies on the observed inverse dependence of W_{ELM} on f_{ELM}. For type I ELMs, using a multi-machine database and a wide range of plasma parameters averaged over multiple ELM events it has been empirically found that [139],

$$\bar{W}_{ELM} = 0.2 W_{plasma} \left(\frac{\bar{\Delta} t_{ELM}}{\tau_E} \right). \tag{8.1}$$

Here, τ_E is the energy confinement time in plasmas with a stored energy W_{plasma} and $\bar{\Delta} t_{ELM}$ is the average period of the ELM cycle ($\bar{\Delta} t_{ELM} = 1/f_{ELM}$). ELM control methods exploit a similar inverse dependence between f_{ELM} and energy loss by increasing the f_{ELM} significantly beyond the natural frequency, leading to smaller ELM energy losses.

As ELM events are repetitive and not periodic, $\bar{\Delta} t_{ELM}$ is customarily estimated

as

$$\bar{\Delta}t_{ELM} = \frac{1}{N}\sum_{i=1}^{N}\Delta t_{ELM_i}. \tag{8.2}$$

Here Δt_{ELM_i} is the time since the previous ELM and is also frequently referred to as the *waiting time* of ELM i. In this work, in contrast to analyzing the relation of the averages, the relation between Δt_{ELM_i} and W_{ELM} for individual ELMs is investigated in a set of JET plasmas with PFCs made of carbon fiber composites (hereafter carbon-wall or CW) and ITER material combination (Be and W) (hereafter ITER-like wall or ILW). In an earlier investigation, Webster *et al.* [108] observed that the inverse dependence between W_{ELM} and f_{ELM} is not obeyed by individual ELMs for Δt_{ELM} greater than $20ms$. However, their analysis was restricted to a set of 2 T, 2 MA ILW plasmas from the JET tokamak. In this work, the analyzed plasmas are selected to cover a wide range of plasma parameters in JET. The aim is to show that an inversely linear relation similar to (8.1) is obeyed in some plasmas, but not all. The correlation between Δt_{ELM} and W_{ELM} is seen to vary in CW discharges and it is usually low in ILW plasmas, except when nitrogen is seeded into the plasma. This is further investigated by examining the relation between ELM durations (τ_{ELM}) and W_{ELM}, as well as the correlation between energies of consecutive ELMs. This includes a comparative analysis between ILW and CW plasmas. A weak or no relation between waiting times and ELM energies could adversely affect the potential of ELM control methods. Therefore, the present work also aims to emphasize the importance of considering the probability distribution of stochastic plasma quantities (in this case Δt_{ELM} and W_{ELM}), as it contains more information compared to a mere average.

Finally, with the aim to locate regions of the machine operational space where ELM control would have a reliable effect on ELM energies, a regression analysis is performed of the correlation between Δt_{ELM} and W_{ELM} on several global plasma parameters.

The structure of the paper is as follows. In section 8.2, we describe the dataset as well as the estimation of the ELM characteristics Δt_{ELM}, W_{ELM} and τ_{ELM}. We also present the statistical tools that are used to assess the strength of the relation between the various parameters of interest. In section 8.3, first the relation between the average quantities is investigated, followed by a similar analysis on the same

		CW	ILW	ILW with N_2 seeding
No. of discharges		20	32	6
Toroidal field	$B_t(T)$	1.6 - 3.0	1.3 - 2.7	2.65 - 2.7
Plasma current	$I_p(MA)$	1.5 - 3.0	1.3 - 2.5	2.5
Line-integrated edge density	$n_e(10^{19}m^{-2})$	3.2 - 9.9	1.9 - 7.4	5.4 - 7.4
Input power = $P_{ohmic} + P_{NBI}$	$P_{input}(MW)$	8.1 - 22	6.9 - 19	16 - 19
Main gas (D_2) flow rate	$\Gamma_{D_2}(10^{22}s^{-1})$	0.0 - 7.5	0.52 - 4.0	1.3 - 3.7
(N_2) flow rate	$\Gamma_{N_2}(10^{22}s^{-1})$	-	-	0.76 - 2.8
Average triangularity	δ_{avg}	0.27 - 0.43	0.27 - 0.41	0.27 - 0.39
Edge safety factor	q_{95}	2.8 - 3.6	3.1 - 6.1	3.4
Beta normalized	β_N	1.6 - 2.4	0.92 -2.0	1.2 - 1.7

Table 8.1: *Range of some key global plasma parameters for the JET ILW, JET CW and the six N_2-seeded JET ILW plasmas from JET ELM-DBII.*

quantities for individual ELMs in a specific discharge. We then study the picture that emerges when all individual ELMs from our database are analyzed together. This is followed by regression analysis of the correlation between waiting times and energy losses, as a function of machine parameters in section 8.4. Finally, in section 8.5 we analyze W_{ELM} of consecutive ELMs before concluding the work in section 8.6.

8.2 Database and methods for correlation analysis

8.2.1 Plasma scenario

For this investigation, an intermediate-size database of 20 CW and 32 ILW JET plasmas has been compiled. We call this database "JET ELMy database (DBII)", henceforth referred as JET ELM-DBII. The database is presented in appendix A.3. The dataset has been selected with a view on encompassing a relatively wide range of plasma and engineering parameters. Each selected discharge has a steady period of H-mode with regular type I ELMs and the analysis has been restricted to time intervals where plasma conditions are quasi-stationary. To ensure quasi-stationarity,

it has been regarded essential that in the analyzed time interval the plasmas have approximately constant gas fueling, input power, edge density and β_N. The size of the current database has somewhat been restricted by the necessary level of manual intervention for extracting data and in part due to the required availability of signals with a sufficient temporal resolution. However, the current size of the database is adequate for the analysis carried out in this work.

With the replacement of CW in JET by the ILW in 2010, it has been observed that the first wall material appears to have had an effect on both the plasma confinement and pedestal properties [171][172]. Up until now, the JET-ILW standard baseline scenario has not routinely achieved a confinement factor of $H_{98} = 1$ both in low and high-triangularity scenarios. The degraded confinement in JET ILW plasmas is a result of a lower pedestal pressure mainly due to a pedestal temperature approximately 20-30 percent lower than in JET CW. Pedestal density on the other hand is comparable among JET CW and JET ILW plasmas. In JET ILW a pedestal pressure comparable to baseline JET CW has only been achieved in high-triangularity experiments with nitrogen (N_2) seeding [172][35]. In the current work, 6 ILW plasmas with N_2 seeding are also included in the dataset, making the total number of analyzed ILW plasmas 38. The range of a number of important engineering parameters in the database is given in table 8.1.

8.2.2 ELM detection and energy loss estimation

A robust threshold-based algorithm has been developed for estimating ELM temporal properties, that is Δt_{ELM} and τ_{ELM}. The algorithm examines Balmer alpha radiation from Deuterium (D_α) for the CW plasmas and Beryllium II (527 nm) radiation for ILW plasmas at JET's inner divertor. The algorithm uses the sharp spikes in D_α/Be II radiation for detecting ELMs. This is preceded by a smoothing process of the time traces and is followed by a threshold-based detection of ELM start and end times. The estimation of Δt_{ELM} and τ_{ELM} is illustrated in figure 8.1. The ELM energy loss has been estimated from the high-resolution time-resolved measurement of the equilibrium stored energy (W_{MHD}). W_{MHD} is calculated by plasma boundary and pressure reconstruction, assuming constant pressure on magnetic surfaces. The W_{MHD} time trace is synchronized to individual ELMs and W_{ELM} is estimated as the maximum loss in energy in a small time window

around an ELM event. This is illustrated in figure 8.2. The time window (delimited by t_a and t_b) is chosen dynamically, with t_a taken as 3/4 of the time till the next ELM and t_b taken as 1/3 of the time since the last ELM. Dynamic selection of the time window compensates for the varying timescales of ELM energy loss between JET CW and JET ILW plasmas [34]. Furthermore, in order to offset inaccuracy arising due to eddy currents in the vacuum vessel and small radial plasma motion following an ELM, a time interval of 3 ms has been allowed after an ELM in which the data is not used for energy loss estimation.

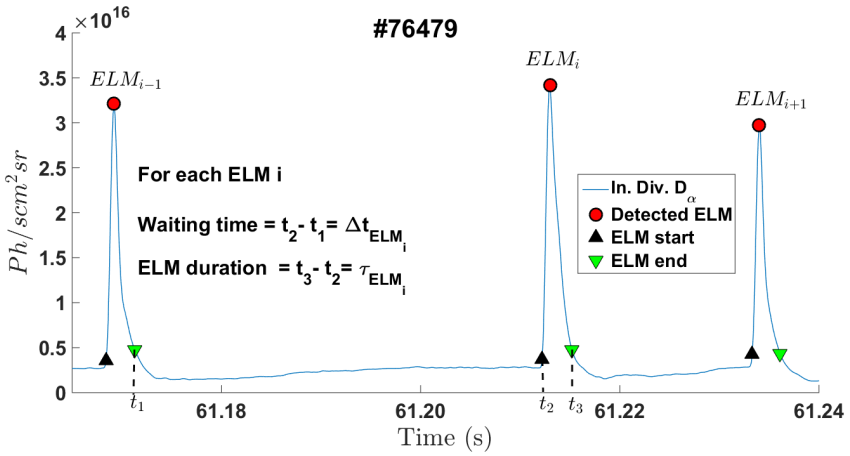


Figure 8.1: *Illustration of the extraction of ELM waiting times (Δt_{ELM}) and ELM durations (τ_{ELM}) from a time trace of D_α radiation at JET's inner divertor.*

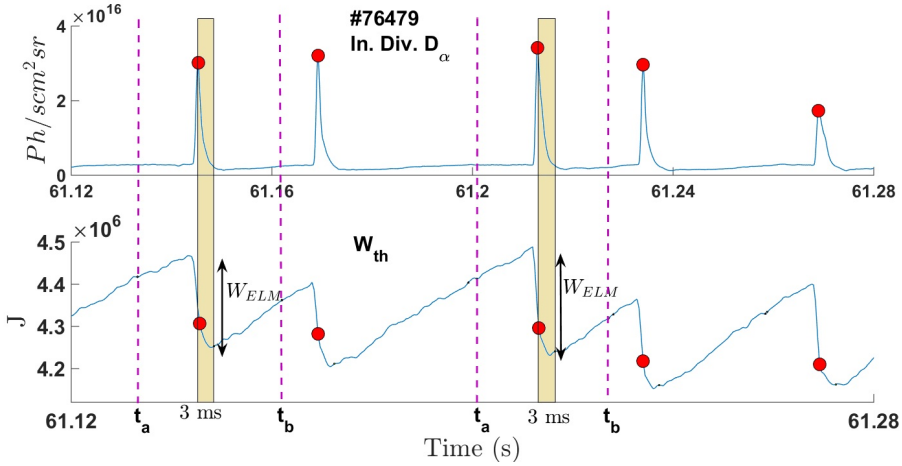


Figure 8.2: *Illustration of ELM energy loss (W_{ELM}) estimation from the equilibrium stored energy (W_{MHD}), synchronized to the time trace of D_α radiation at JET's inner divertor.*

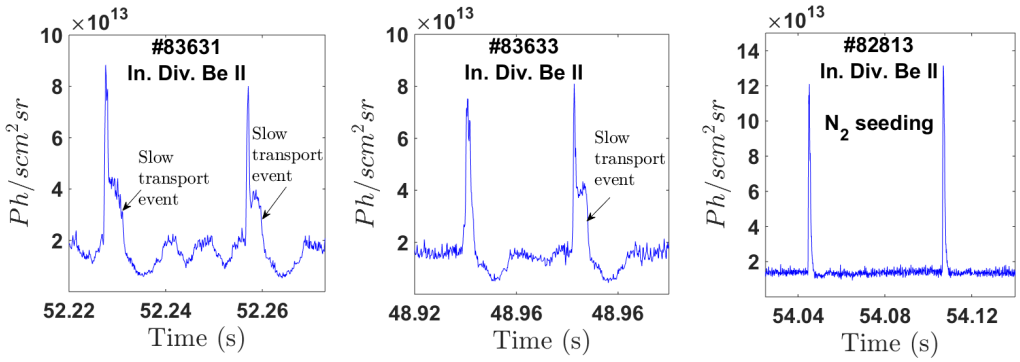


Figure 8.3: *Temporal signature of pure ELMs and ELMs followed by a slow transport event (STE) in three typical JET ILW plasmas. The N_2-seeded plasmas, like CW plasmas, have narrower ELMs and no slow transport events.*

8.2.3 ELM duration and slow transport events

JET ITER-like wall ELMs are sometimes followed by an extended collapse phase, called the *slow transport event* (STE) [34]. These STEs are analogous to the second phase of ELM collapse observed at ASDEX Upgrade (AUG) [35]. The typical temporal signature of an STE is shown in figure 8.3. ELMs accompanied by an STE have longer time scales of temperature and density collapse and result in higher total energy loss of the plasma than the losses produced by ELMs alone. We first studied the variation of the energy released by an ELM, averaged over all ELM events in a single discharge, in terms of the fraction of STEs. The latter is defined as

$$f_{STE} = \frac{N_{(ELM+STE)}}{N_{ELM} + N_{(ELM+STE)}}, \tag{8.3}$$

where $N_{(ELM+STE)}$ is the number of ELMs accompanied by a slow transport event and N_{ELM} is the number of ELMs that are not followed by an STE phase, hereafter referred to as "pure" ELMs. The ELM energy loss averaged over a single discharge, during stationary conditions, is denoted as $\bar{W}_{ELM}$ and we also consider its ratio w.r.t. $\bar{W}_{tot}$, i.e. the total stored equilibrium energy in the plasma, also averaged over the entire stationary phase of each discharge that has been investigated. The variation of $\bar{W}_{ELM}$ and $\bar{W}_{ELM}/\bar{W}_{tot}$ with the fraction of STEs (f_{STE}) for all plasma pulses is plotted in figure 8.4. In this work, we have divided JET ILW plasmas (N discharges) into three broad categories: those with a high fraction of STEs ($f_{STE} \geq 50\%, N = 4$), medium fraction of STEs ($10\% \leq f_{STE} < 50\%, N = 24$) and those with very few or no STEs ($f_{STE} < 10\%, N = 4$). From figure 8.4, a clear (linear) increase can be

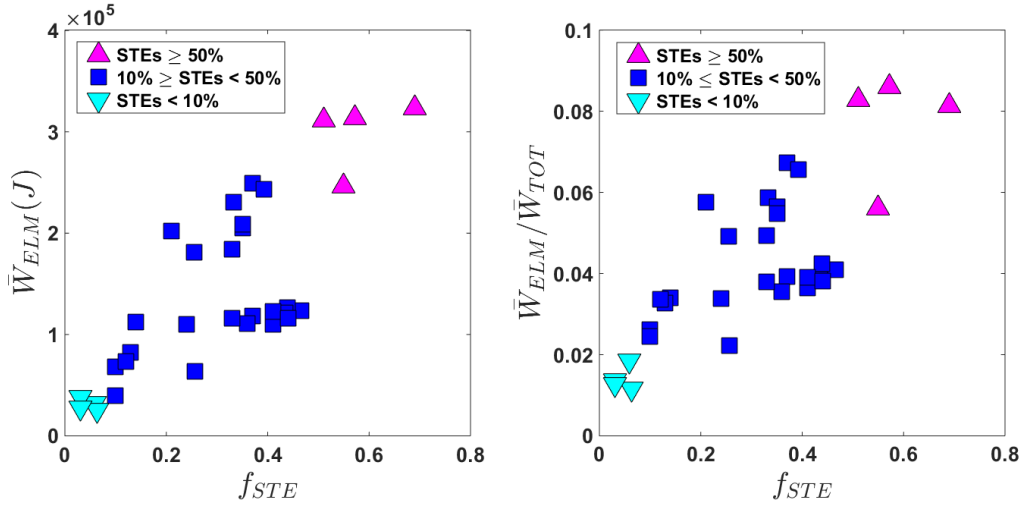


Figure 8.4: *Variation of the mean ELM energy loss ($\bar{W}_{ELM}$) and mean relative ELM energy loss ($\bar{W}_{ELM}/\bar{W}_{tot}$) with the fraction of slow transport events (f_{STE}) in JET ILW plasmas.*

noticed of $\bar{W}_{ELM}$ with the fraction of STEs in a plasma. A very similar conclusion is true for the relative energy loss $\bar{W}_{ELM}/\bar{W}_{tot}$, which shows that an increased energy loss is due to a higher fraction of STEs. This is in accordance with recent studies wherein it was seen that the STEs carry a significant proportion of the energy of the total ELM event [34]. STEs are absent in the JET CW database analyzed in this work. Furthermore, they disappear in N_2-seeded ILW JET plasmas [34], as does the second part of the ELM collapse in AUG plasmas [35]. JET ILW ELMs, compared to JET CW plasmas have larger ELM durations (τ_{ELM}). This too, in a large part, is due to the existence of STEs in ILW plasmas. The average duration $\bar{\tau}_{ELM}$ of all ELM events during a period of stationary plasma conditions, for the plasmas analyzed in this work, are listed in table 8.2. N_2-seeded ILW plasmas and ILW plasmas with low f_{STE} have $\bar{\tau}_{ELM}$ similar to CW plasmas. ILW plasmas with high f_{STE} exhibit $\bar{\tau}_{ELM}$ about three times larger than the $\bar{\tau}_{ELM}$ of CW plasmas.

Table 8.2: *Typical ELM durations (mean ($\bar{\tau}_{ELM}$) and standard deviation ($std(\tau_{ELM})$)) for unseeded JET ILW plasmas (varying degrees of slow transport events), N_2-seeded JET ILW plasmas and JET CW plasmas.*

	$\bar{\tau}_{ELM}(ms)$	$std(\tau_{ELM})(ms)$
ILW		
$f_{STE} \geq 50\%$	7.1	3.8
$10\% \leq f_{STE} < 50\%$	3.4	2.2
$f_{STE} < 10\%$	2.7	0.8
N_2-seeded	2.5	0.8
CW	2.6	1.2

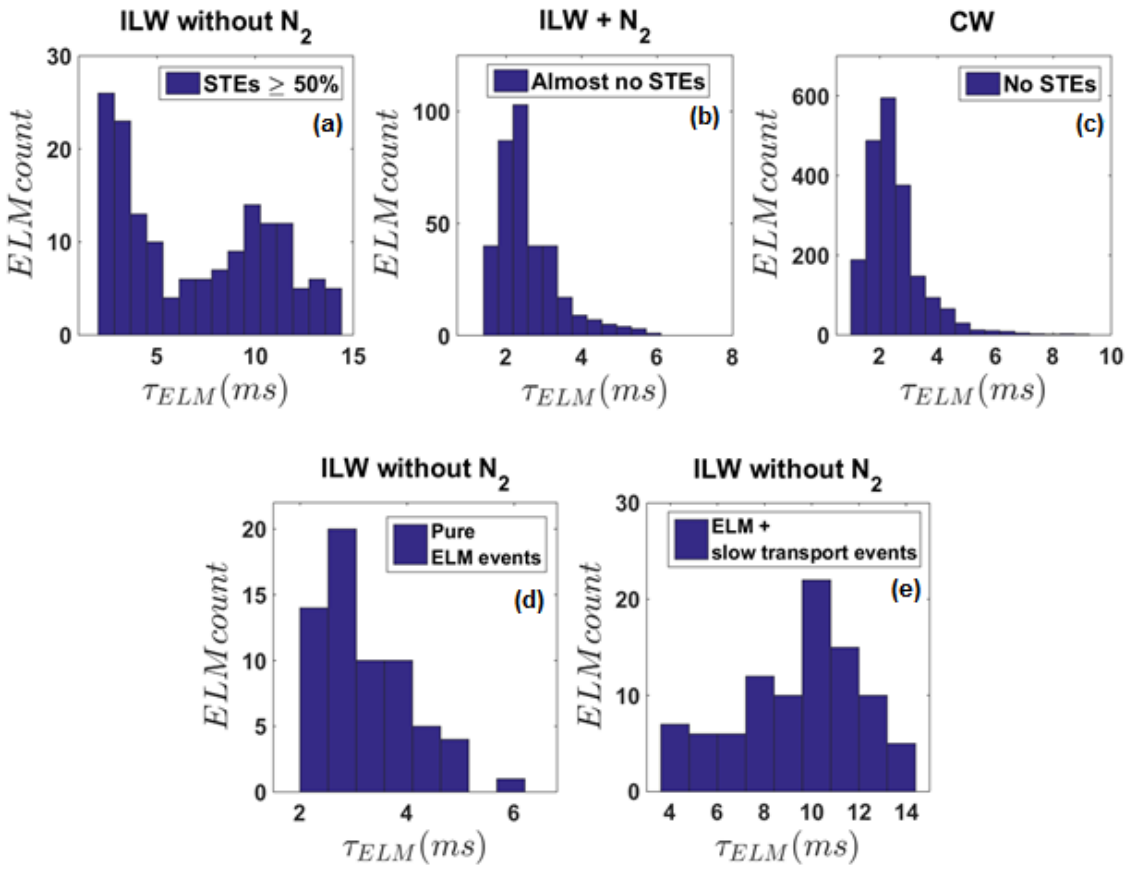


Figure 8.5: *Distribution of ELM durations for various subsets of JET plasmas investigated in this work. In each panel, the vertical axis shows the number of ELM events. (a) Unseeded ILW plasmas with a high f_{STE}, (b) N_2-seeded ILW plasmas, (c) CW plasmas, (d) Pure ELMs from high f_{STE} unseeded ILW plasmas, (e) ELMs followed by STEs from high f_{STE} unseeded ILW plasmas.*

An investigation into the distribution of τ_{ELM} yields that the non-seeded JET ILW plasmas (high f_{STE}) have a distribution of τ_{ELM} which is distinctly different from N_2-seeded JET ILW plasmas and JET CW plasmas. The latter two cases exhibit similar distributions for τ_{ELM}. Figure 8.5 (a)-(c) present the distribution of τ_{ELM} for non-seeded JET ILW plasmas (high f_{STE}), N_2-seeded JET ILW plasmas and JET CW plasmas. The distribution of τ_{ELM} for non-seeded JET ILW plasmas (high f_{STE}) is bimodal (two local maxima). The bimodal distribution arises as a mixture of two underlying unimodal distributions emerging from collapses due to pure ELMs and collapses followed by STEs. We performed a manual separation of pure ELM events from the cases with STEs, and the corresponding unimodal distributions are shown in figure 8.5(d) and (e), respectively. The pure ELMs have a duration τ_{ELM} that is typically less than about 5 ms, while the ELMs with STEs can last up to 14 ms. The distribution of τ_{ELM} for pure ELMs in high f_{STE} ILW plasmas

(figure 8.5(d)) appear similar to the distribution of τ_{ELM} for N_2-seeded JET ILW plasmas (figure 8.5(b)) and JET CW plasmas (figure 8.5(c)). These distributions are visibly non-Gaussian with a strong positive skew and we verified that a similar degree of skewness also exists in the distribution of ELM durations from individual discharges. From the physical point of view it means that, in our data set, pure ELMs with durations longer than 4 - 5 ms are relatively rare, compared to the prevailing duration of about 2.5 ms. From the statistical point of view, characterization of skewed distributions necessitates additional metrics such as median and mode. The means and standard deviations alongside medians, and skewness estimates for each distribution are summarized in table 8.3. Here, the skewness was estimated not from

Table 8.3: *Summary (mean ($\bar{\tau}_{ELM}$), standard deviation ($std(\tau_{ELM})$), median ($\tilde{\tau}_{ELM}$) and skewness) for the distributions of ELM durations extracted from the JET discharges investigated in this work.*

JET plasmas		$\bar{\tau}_{ELM}$ (ms)	$std(\tau_{ELM})$ (ms)	$\tilde{\tau}_{ELM}$ (ms)	Skewness
ILW plasmas	Pure ELMs	3.2	0.87	3.0	0.23
$f_{STE} \geq 50\%$	ELMs + STEs	9.6	2.5	9.8	0.08
N_2-seeded ILW plasmas		2.5	0.81	2.3	0.25
CW plasmas		2.6	1.2	2.3	0.25

the third-order moment of the distribution (which typically requires a lot of data points), but by dividing the difference between mean and median with standard deviation. For gaining an interesting insight into skewness estimation, the reader may refer to [173]. Contrary to pure ELM events, the distribution of τ_{ELM} for ELMs followed by STEs in high f_{STE} JET ILW plasmas (figure 8.5(e)) follows a more symmetric distribution.

8.2.4 Tools for relation analysis

For analyzing the relation between ELM waiting times and energy losses, as a first step we use scatter graphs to get a qualitative impression. Furthermore, in order to quantify the strength of linear relation between Δt_{ELM} and W_{ELM} for individual ELMs within single discharges, the regular Pearson's product moment correlation coefficient (ρ) is estimated [174] [175]. For two sets of data or random variables X

and Y, this correlation coefficient is defined as,

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y}, \tag{8.4}$$

where *cov* stands for the covariance between the variables, while σ_X and σ_Y are their standard deviations. $\rho_{X,Y}$ takes values in the range $[-1, 1]$; a value of 1 means that X and Y are perfectly linearly correlated, a value of 0 that there is no correlation, while a value of -1 that they are perfectly anti-correlated.

Further statistical inference that we will perform based on ρ includes estimation of confidence intervals, testing the significance of correlations and regressing against a set of global engineering parameters. This is complicated by the in general non-Gaussian distribution of a correlation coefficient. Therefore estimates r of ρ are converted to a z-value, which is known to follow an approximately normal distribution:

$$z \equiv \frac{1}{2} \ln \frac{(1+r)}{(1-r)} = \tanh^{-1}(r). \tag{8.5}$$

The mean of the distribution is the z-value itself, while the standard deviation does not depend on r and can be approximated by $\sigma_z = 1/\sqrt{n-1}$, where n is the number of data points. In addition, we use an alternative measure of relation, in order to capture any possible nonlinear relation between the variables under investigation. This is Spearman's rank correlation coefficient r_s, which measures monotonic dependence between X and Y:

$$r_s = 1 - \frac{6 \sum_{i=1}^{n} (X_i - Y_i)^2}{n(n^2-1)}, \tag{8.6}$$

where X_i denotes the rank of the value X_i in the ordered series of values of the variable X. r_s is a nonparametric measure of dependence and is much less sensitive to outliers. Similar to r, r_s is in the interval [-1,1] and $r_s = 0$ implies no monotonic dependence.

Finally, partial correlation is also used when treating ELMs from different plasmas at the same time. Partial correlation measures the degree of association between two random variables while correcting for the effect of another variable, or several other variables, on this relation. The partial correlation of X and Y, adjusted for

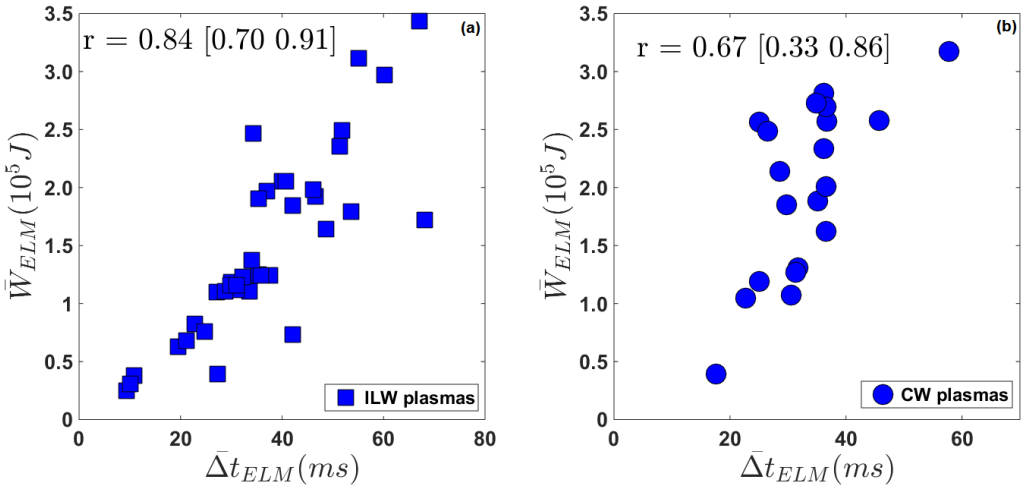


Figure 8.6: *Scatter graphs between $\bar{W}_{ELM}$ and $\bar{\Delta}t_{ELM}$ for (a) JET ILW plasmas, (b) JET CW plasmas. Estimates for the Pearson correlation coefficient (r) are indicated, together with the 95% confidence interval.*

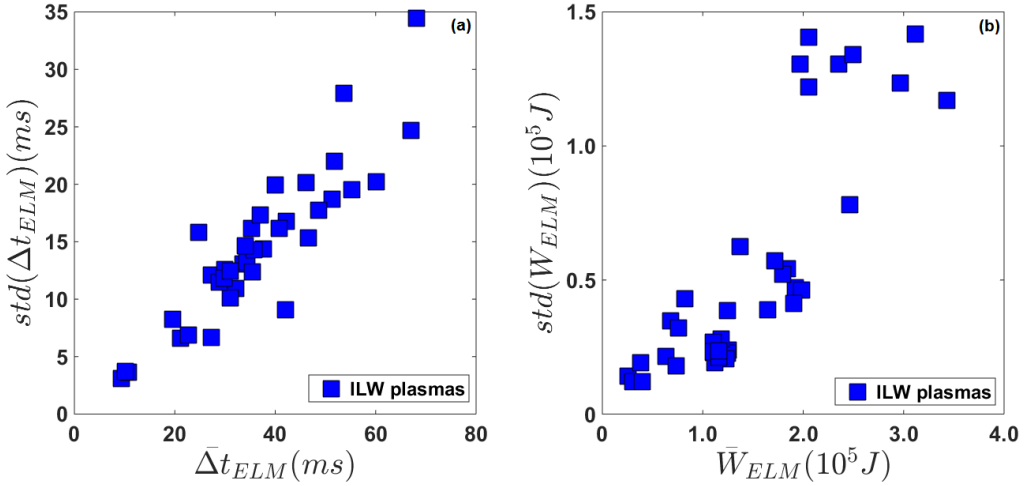


Figure 8.7: *Scatter graphs between mean and standard deviation of (a) Δt_{ELM} and (b) W_{ELM}, for the JET ILW plasmas.*

Z is:

$$\rho_{XYZ} = \frac{\rho_{XY} - \rho_{XZ}\rho_{YZ}}{\sqrt{(1-\rho_{XZ}^2)(1-\rho_{YZ}^2)}}. \tag{8.7}$$

Partial correlation can also be computed for Spearman's rank correlation coefficient.

8.3 Analysis of the relation between ELM properties

The relation between W_{ELM} and Δt_{ELM}, averaged over all ELMs in a single discharge, is shown in figure 8.6(a) and (b) for ILW and CW plasmas, respectively. In agreement with the findings in [139], there is a strongly positive correlation

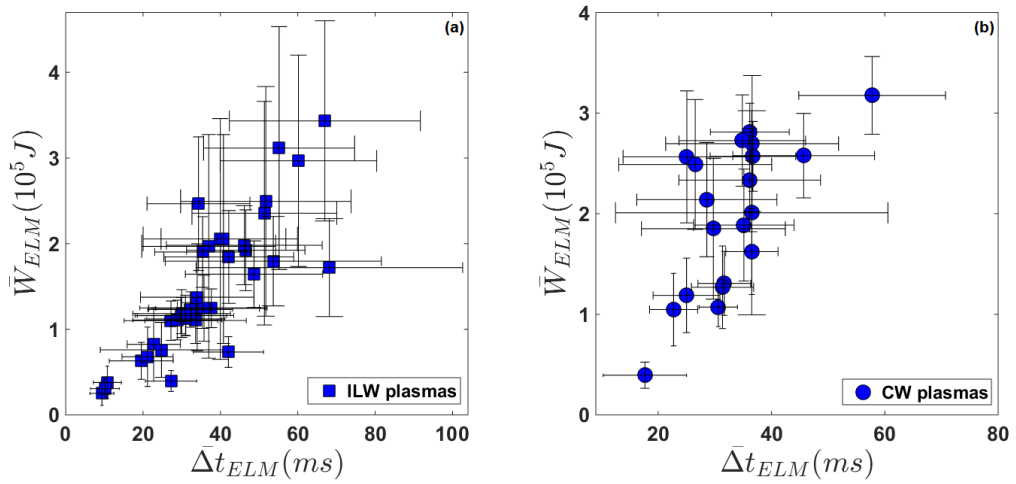


Figure 8.8: *Scatter graphs between $\bar{W}_{ELM}$ and $\bar{\Delta}t_{ELM}$, including the error bars specified by a single standard deviation, for (a) JET ILW plasmas, (b) JET CW plasmas.*

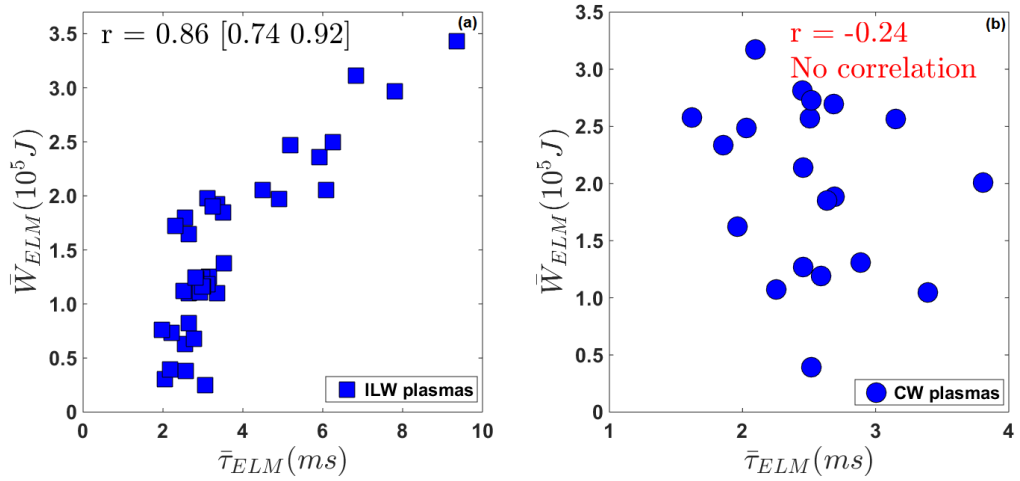


Figure 8.9: *Scatter graphs between $\bar{W}_{ELM}$ and $\bar{\tau}_{ELM}$ for (a) JET ILW plasmas, (b) JET CW plasmas. Estimates for the Pearson correlation coefficient (r) are indicated, together with the 95% confidence interval. CW plasmas, in contrast to ILW plasmas, fail to reject the null hypothesis of no correlation at 5% significance level.*

between W_{ELM} and Δt_{ELM} for ILW plasmas as well as for CW plasmas. However, ELM control is targeted at influencing the energy loss of individual ELMs. Thus, basing the mitigation strategy on the relation between the average properties of different plasmas can possibly be an oversimplification. Furthermore, the relation presented in [139] does not take into account the uncertainty on W_{ELM} and Δt_{ELM}. Nevertheless, it can be observed from figure 8.7 that the standard deviation of W_{ELM} and Δt_{ELM} is substantial and increases roughly linearly with the mean value. A straightforward extrapolation based on figure 8.7(b) would suggest 7 - 10 MJ of standard deviation around an absolute W_{ELM} of 20 - 30 MJ at ITER.

In general, the probability distributions of ELM properties contain comprehensive

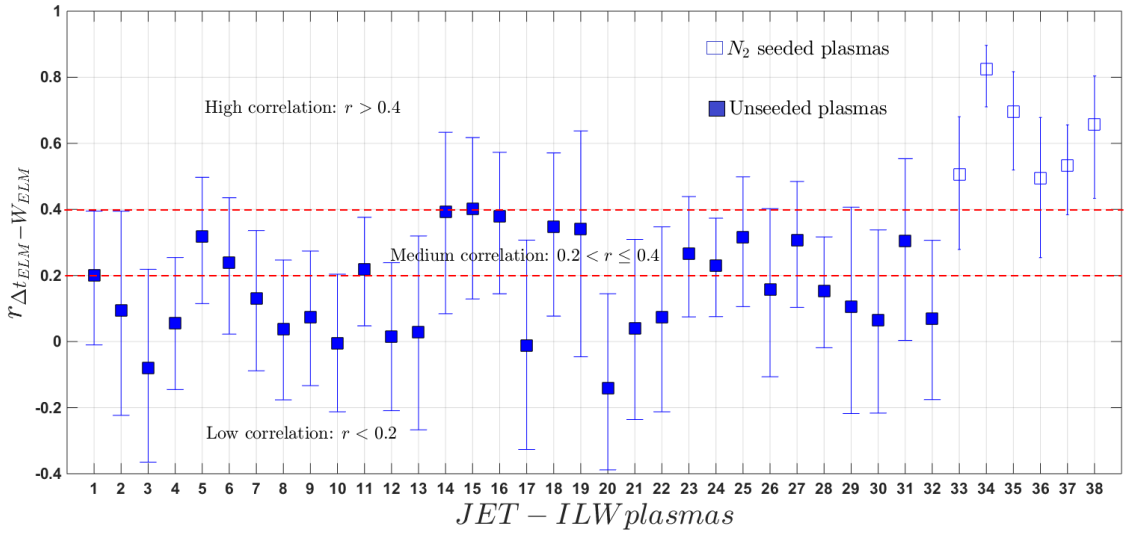


Figure 8.10: *Estimates of linear correlation between W_{ELM} and Δt_{ELM} for individual ELMs in JET ILW plasmas. 95% confidence intervals are also indicated. Discharges indexed 33 to 38 are N_2-seeded plasmas.*

information about their variability [157][89][176] and therefore studying their statistical correlation properties will yield a better insight into the strength of any existing relations. Figure 8.8 is essentially a reproduction of figure 8.6, with the addition of the error bars indicating a single standard deviation. The strongly linear relations depicted in figure 8.6 appear to be less clear with the inclusion of standard deviations in figure 8.8. Hence, as will be shown below, the effect of the spread in W_{ELM} and Δt_{ELM} within each plasma is better quantified by studying the relation between W_{ELM} and Δt_{ELM} for individual ELMs in a discharge. Furthermore, the relation between W_{ELM} and τ_{ELM} for ILW and CW plasmas is shown in figure 8.9. The correlation is clearly different in the two cases: ILW plasmas exhibit a strongly positive correlation, whereas CW plasmas, failing to reject the null hypothesis of zero correlation at 5 percent significance level, effectively show no correlation.

8.3.1 Properties of individual ELMs

After studying the ELM properties averaged over a window of stationary plasma conditions, we now concentrate on relations between the properties of the individual ELMs. Estimates of the correlation between W_{ELM} and Δt_{ELM} ($r_{\Delta t_{ELM}-W_{ELM}}$), along with 95% confidence intervals are presented in figure 8.10 and figure 8.11 for individual ELMs in JET ILW and JET CW plasmas, respectively. Despite $\bar{W}_{ELM}$ and $\bar{\Delta t}_{ELM}$ conforming to the expected inverse dependence between W_{ELM} and

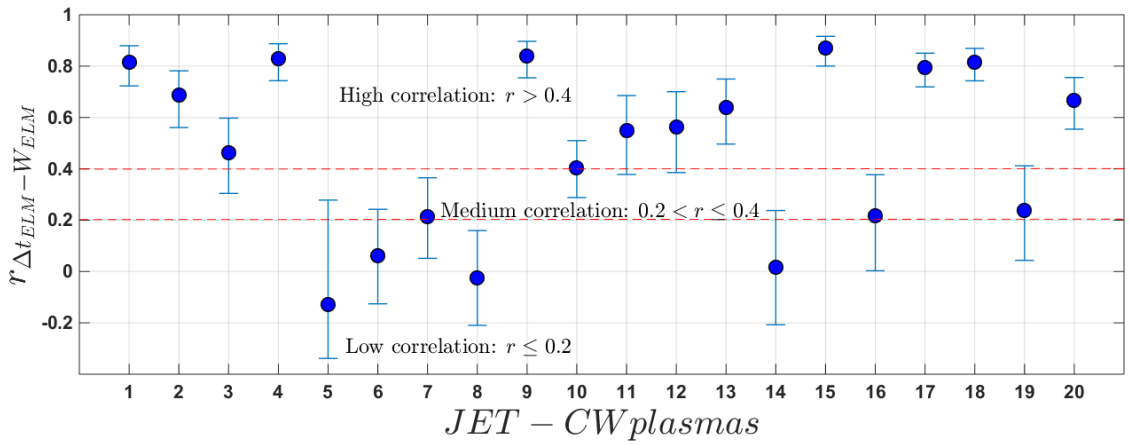


Figure 8.11: *Estimates of linear correlation between W_{ELM} and τ_{ELM} for individual ELMs in JET CW plasmas. 95% confidence intervals are also indicated.*

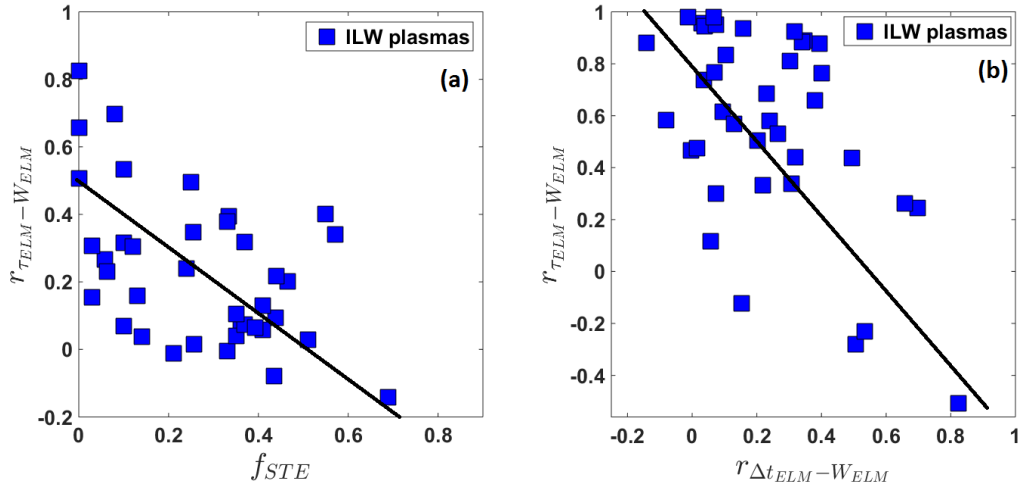


Figure 8.12: *Variation of linear correlation between W_{ELM} and Δt_{ELM} ($r_{(\Delta t_{ELM})-W_{ELM})}$) for individual ELMs in JET ILW plasmas. (a) With the fraction of slow transport events (f_{STE}) and (b) with the linear correlation between W_{ELM} and τ_{ELM} ($r_{(\tau_{ELM}-W_{ELM})}$) for individual ELMs in JET ILW plasmas.*

f_{ELM}, the correlation between W_{ELM} and Δt_{ELM} for individual ELMs varies from being strongly correlated for certain plasmas to being uncorrelated for others. This is observed in both CW as well as ILW plasmas. Compared to ILW discharges, CW plasmas on the whole have higher correlation between W_{ELM} and Δt_{ELM} for individual ELMs, with 12 out of the 20 (60%) analyzed plasmas exhibiting high correlation ($r > 0.40$) and 4 out of the 20 (20%) analyzed plasmas demonstrating no correlation ($r \leq 0.20$). On the other hand, out of the 38 ILW plasmas, only the 6 (16%) N_2-seeded plasmas exhibit high correlation ($r > 0.40$), whereas 19 (50%) plasmas show no correlation and 13 (34%) have a medium correlation.

The underlying processes causing W_{ELM} and Δt_{ELM} to exhibit varying degrees of

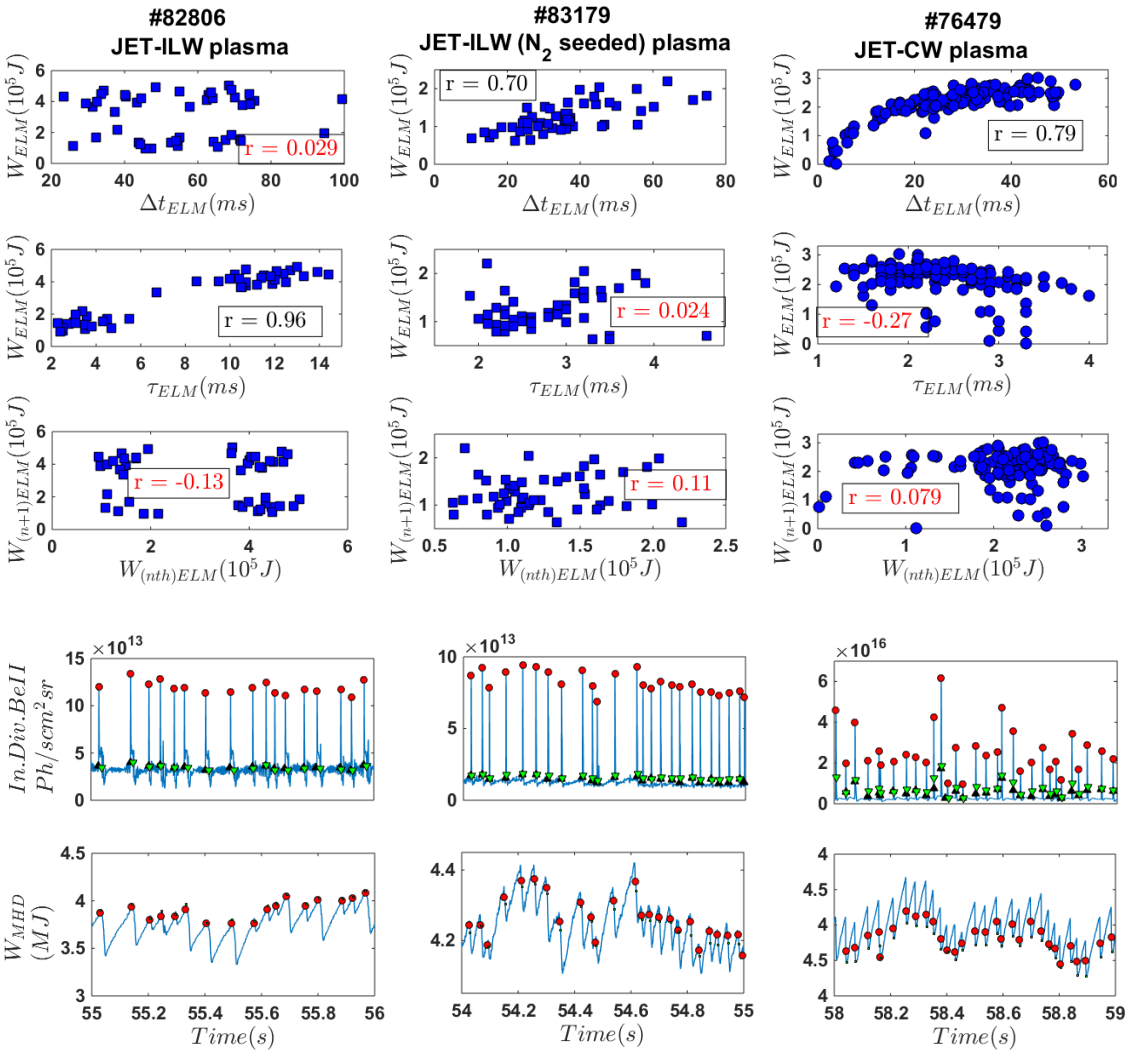


Figure 8.13: *Scatter plot between W_{ELM} and Δt_{ELM}, W_{ELM} and τ_{ELM} and $W_{(nth)ELM}$ and $W_{(n+1)ELM}$ for JET pulse #82806 (unseeded JET ILW plasma (STEs > 50%)), #83179 (N_2-seeded JET ILW plasma) and #76479 (JET CW plasma). Estimates of r for each scatter plot are also specified. r estimates that fail to reject the hypothesis of no correlation at 5% significance level are indicated in color red. Also given are time traces of Be II radiation from the inner divertor (ILW plasmas), D_α from the inner divertor (CW plasma) and the equilibrium stored energy (W_{MHD}).*

correlation could be one or several of the following. The size of W_{ELM} is controlled by the pedestal parameters, i.e. the density and temperature inside the pedestal before the ELM crash [103][107]. A multi-machine study performed on ASDEX, DIII-D, JT60U and JET CW has established that the relative ELM energy losses scale with the inverse of pedestal collisionality [103]. Other key parameters that have an important effect on W_{ELM} are the pedestal width [130], plasma rotation [177] and the plasma shape [178]. On the other hand, Δt_{ELM} is a consequence of

the various timescales involved in the recovery of the pedestal to its pre-ELM state following the ELM crash. The pedestal recovery time can be potentially modified by enhanced losses in the inter-ELM period, either by increased bulk radiation or by an increased level of density and magnetic fluctuations. W_{ELM}, being determined primarily by the pre-ELM pedestal plasma parameters, is likely to remain unaffected by the inter-ELM processes that can potentially modify Δt_{ELM}. Furthermore, the peeling-ballooning model, which is a leading candidate for explaining ELM onset, fails to explain the phase of saturated gradients without ELMs [179]. In medium-sized tokamaks at low edge temperature, the bootstrap current seems to be fully developed for a relatively long time interval before an ELM crash. It is reasonable to assume that, after the pedestal has recovered, an additional increase in Δt_{ELM} will not lead to an additional increase in W_{ELM}. Finally, figure 8.12 suggests that, in the case of the ILW plasmas, the correlation between W_{ELM} and Δt_{ELM} for individual ELMs varies inversely with f_{STE}. Hence, the presence of the STEs appears to be at least partly responsible for the observed reduction in correlation between ELM waiting times and energies in ILW plasmas.

Furthermore, we note that for ILW plasmas there is a weakly inverse relation between the correlation among W_{ELM} and Δt_{ELM} and the correlation among τ_{ELM} and W_{ELM}. It can be seen from figure 8.12 that plasmas with high f_{STE} exhibit no correlation between W_{ELM} and Δt_{ELM} and consequently a very high correlation between τ_{ELM} and W_{ELM}. As an illustration, scatter plots between W_{ELM} and Δt_{ELM} and W_{ELM} and τ_{ELM} for three representation plasmas are given in figure 8.13. On the one hand, non-seeded JET-ILW plasma #82806 with $f_{STE} \geq 0.5$ exhibits a very high correlation between W_{ELM} and τ_{ELM} and no correlation between W_{ELM} and Δt_{ELM}. On the other hand, N_2-seeded JET-ILW plasma #83179, similar to JET-CW plasma #76479, demonstrates a high correlation between W_{ELM} and Δt_{ELM} and no correlation between W_{ELM} and τ_{ELM}.

8.3.2 Collective properties of individual ELMs in all analyzed plasmas

Next, the collective properties of all ELM events in our JET ILW database are investigated. A scatter diagram between W_{ELM} and Δt_{ELM} for all ELMs (excluding

N_2-seeded plasmas) is shown in figure 8.14(a). Table 8.4 lists the estimates for r and r_s corresponding to the scatter diagram presented in figure 8.14(a). Partial correlations between W_{ELM} and Δt_{ELM}, while controlling for B_t, I_p, P_{input}, n_e, Γ_{D_2} and δ_{avg}, are presented as well. In this case partial correlation is a more realistic measure for assessing the relation between W_{ELM} and Δt_{ELM}, since it takes into account the widely varying global plasma conditions across the data set. It is noteworthy that adjusting for the varied plasma conditions brings a significant reduction in the correlation. Moreover, values of r_s are comparable with r, which confirms the robustness of r estimates.

Furthermore, in order to account for any variation of the standard deviation of the data (heteroscedasticity), which is especially clear in figure 8.14(a) (see also figure 8.7), a scatter diagram between the logarithm of W_{ELM} and Δt_{ELM} for all ELMs in the analyzed ILW plasmas (excluding N_2-seeded plasmas) is shown in figure 8.14(b). Also, on figure 8.14(b), the least-squares line of best fit is indicated and the corresponding regression coefficients are given in table 8.5. The observed linearity in the log-log space is indicative of a power law relation between W_{ELM} and Δt_{ELM}. This implies that the rate of change of W_{ELM} and Δt_{ELM} decreases gradually up to a point beyond which the two quantities become almost independent. This is reaffirmed by the inspection of figure 8.14(a) where there appears to be a saturation of W_{ELM} for Δt_{ELM} greater than 25-30 ms. This is also in agreement with an earlier observation of statistical independence between W_{ELM} with Δt_{ELM} beyond $\Delta t_{ELM} = 20ms$, made by Webster *et al.* [108] for individual ELMs from a set of $2T$, $2MA$ JET ILW plasmas. The point beyond which W_{ELM} becomes independent of Δt_{ELM} is likely to be limited by the pedestal recovery time and the total energy stored in the plasma. In the plasmas considered in this work, though the plasma thermal energy for pure ELMs appears to increase until the next ELM, it is largely recovered to its pre-ELM value in $25(\pm 8)ms$. This suggests a scenario in which the edge pedestal is largely restored in $\approx 25ms$, leading to a significant reduction in the correlation between W_{ELM} for Δt_{ELM} beyond $\Delta t_{ELM} \approx 25ms$. On the other hand, for ELMs followed by STEs, the plasma thermal energy recovers to its pre-ELM+STE value in $90(\pm 10)ms$. Furthermore, it can be estimated that for ILW ELMs a reduction of Δt_{ELM} from 25-30 ms (beyond which W_{ELM} and Δt_{ELM} are very weakly correlated) to 10 ms reduces W_{ELM} by $\approx 60\%$. On the other hand,

a reduction of Δt_{ELM} from 50-60 ms to 25-30 ms, reduces W_{ELM} by $\approx 40\%$. This suggests that if ELMs are consistently paced at 10 ms, W_{ELM} can be reduced by $\approx 60 - 70\%$.

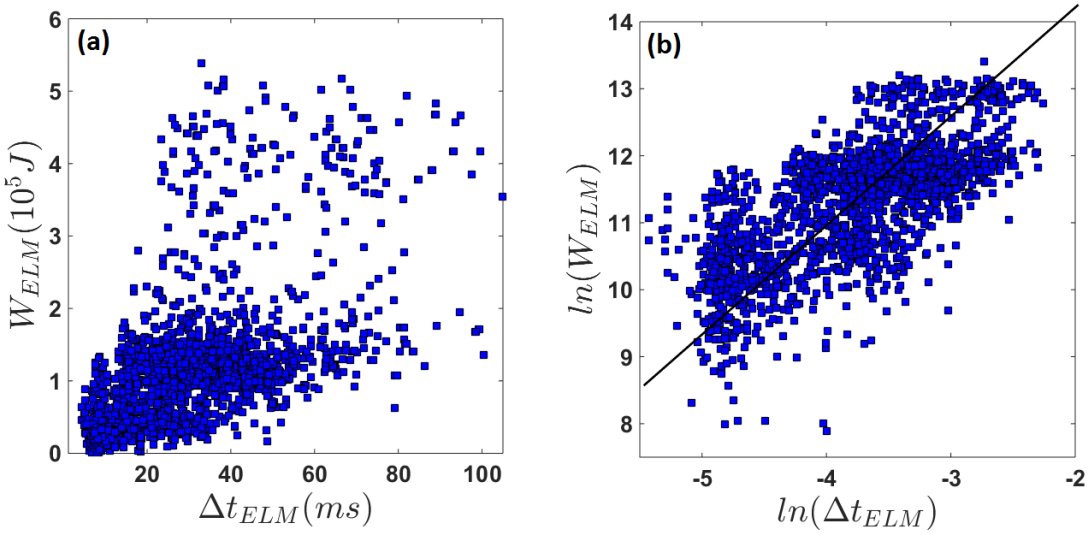


Figure 8.14: *Scatter graph between (a) W_{ELM} and Δt_{ELM}, (b) Logarithm of W_{ELM} and Δt_{ELM} for all ELMs in JET ILW plasmas. The least-squares line of best fit to the logarithm of W_{ELM} and Δt_{ELM} is also shown.*

Table 8.4: *Estimates of regular and partial correlations, based on Pearson (r) and Spearman (r_s) coefficients, between W_{ELM} and Δt_{ELM} for all ELMs in the JET ILW plasmas. The partial correlations control for B_t, I_p, P_{input}, n_e, Γ_{D_2} and δ_{avg}.*

	r	r_s
Regular	0.58	0.65
Partial	0.21	0.26

Table 8.5: *Estimated coefficients and standard errors for the least-squares line of best fit shown in figure 8.14(b). The* model *is $ln(W_{ELM}) = \beta_0 + \beta_1 ln\Delta t_{ELM}$.*

β_0	β_1	SE_{β_0}	SE_{β_1}
14.7	0.895	0.071	0.019

8.4 Global dependence of correlation between ELM energy losses and waiting times

Since the success of ELM mitigation depends considerably on a high correlation between W_{ELM} and Δt_{ELM}, we now aim to locate the regions of plasma operational

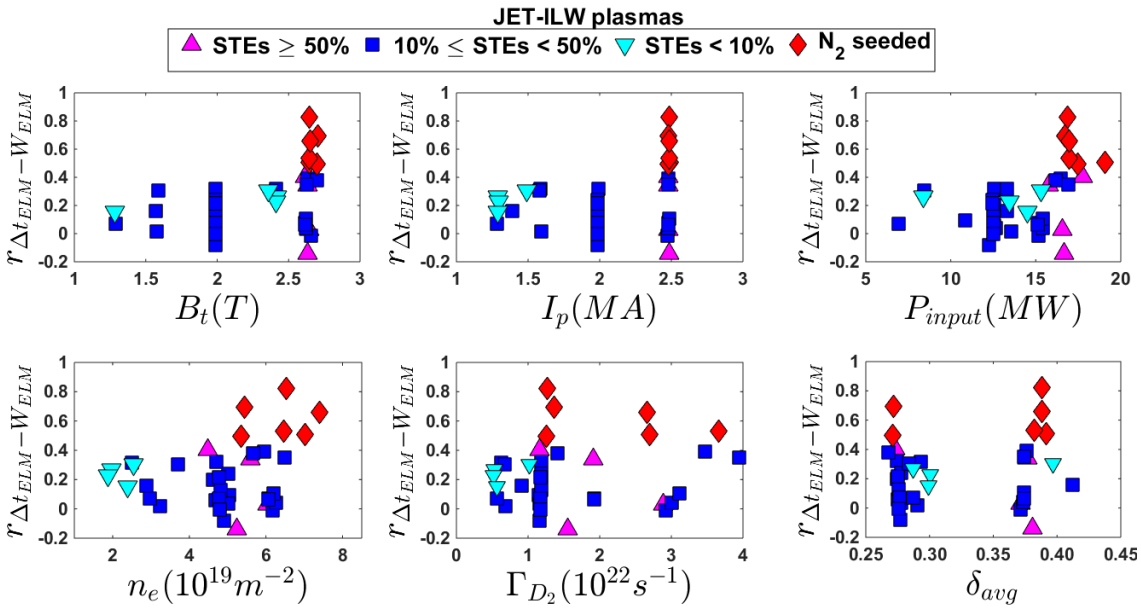


Figure 8.15: *Scatter plots of correlation between W_{ELM} and Δt_{ELM} ($r_{(\Delta t_{ELM}-W_{ELM})}$) and plasma engineering parameters B_t, I_p, P_{input}, n_e, Γ_{D_2} and δ_{avg} for JET ILW plasmas.*

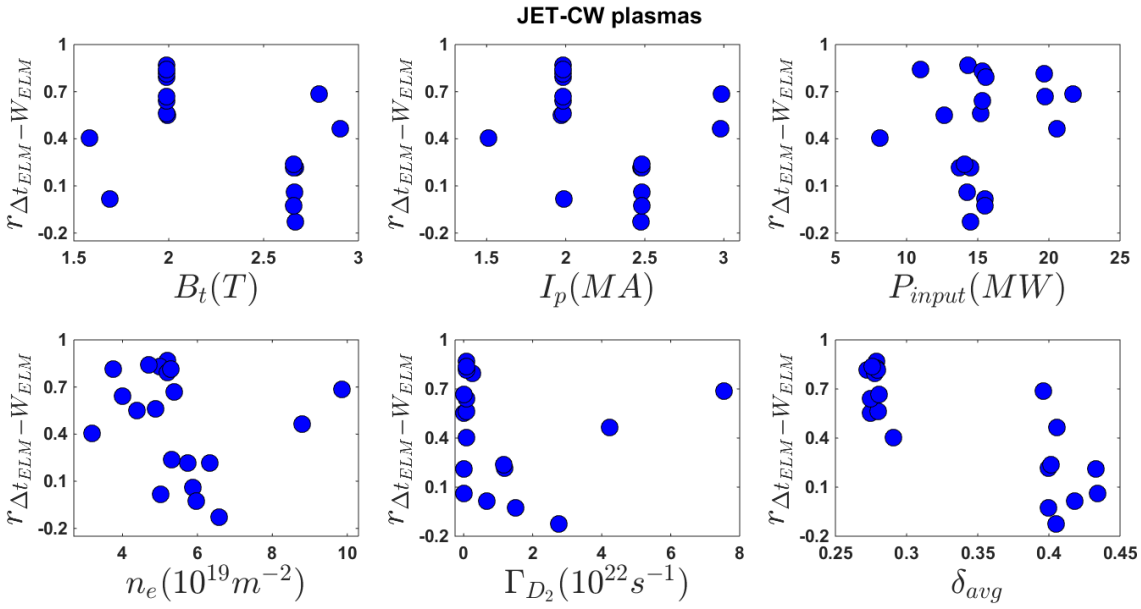


Figure 8.16: *Scatter plots of correlation between W_{ELM} and Δt_{ELM} ($r_{(\Delta t_{ELM}-W_{ELM})}$) and plasma engineering parameters B_t, I_p, P_{input}, n_e, Γ_{D_2} and δ_{avg} for JET CW plasmas.*

space where the corresponding correlation coefficient $r_{(\Delta t_{ELM}-W_{ELM})}$ is large. One approach for studying the dependence of $r_{(\Delta t_{ELM}-W_{ELM})}$ on plasma parameters would be to rely on single parameter scans. In the case of the present work, there are not enough dedicated experiments available to allow such a study. Nevertheless, as a preliminary step, in figure 8.15 and figure 8.16 scatter plots between the plasma engineering parameters B_t, I_p, P_{input}, n_e, Γ_{D_2}, δ_{avg} and the correlation coefficient $r_{(\Delta t_{ELM}-W_{ELM})}$ are provided. It can be observed that individually none of the plasma

Table 8.6: *Least-squares multilinear regression fits (including a cut-off term C) for correlation between W_{ELM} and Δt_{ELM} using global plasma parameters as predictors. The coefficient estimate alongside 95% confidence intervals are presented, together with the root-mean-square error (RMSE) and the coefficient of determination (R^2).*

	CW	ILW	
		Model 1	Model 2
C	1.67 [0.43 2.92]	-0.457 [-1.1 0.15]	0.029 [-0.56 0.62]
$B_t(T)$	-0.982 [-2.4 0.41]	0.0483 [-0.30 0.39]	0.162 [-0.14 0.46]
$I_p(MA)$	1.62 [-0.66 3.9]	0.559 [-0.43 1.5]	0.0791 [-0.69 0.85]
$P_{input}(MW)$	-0.0229 [-0.089 0.043]	0.0119 [-0.036 0.060]	0.0080 [-0.038 0.054]
$n_e(10^{19}m^{-2})$	0.165 [-0.11 0.44]	-0.0259 [-0.24 0.19]	-0.0486 [-0.25 0.15]
$\Gamma_{D_2}(10^{22}s^{-1})$	**-0.113** [-0.26 0.039]	**-0.114** [-0.24 0.012]	**-0.0422** [-0.17 0.084]
δ_{avg}	**-8.54** [-12 -5.4]	**-0.313** [-2.2 1.5]	**-0.618** [-2.3 1.1]
f_{STE}	—-	**-1.19** [-1.7 -0.65]	—-
$\Gamma_{N_2}(10^{22}s^{-1})$	—-	—-	**0.269** [0.16 0.38]
RMSE(%)	23.4	18.3	17.4
R^2	0.83	0.64	0.67

engineering parameters discriminate well between plasmas with a high, medium or zero $r_{(\Delta t_{ELM}-W_{ELM})}$. As a next step, regression analysis is used for quantifying the effect of plasma parameters on $r_{(\Delta t_{ELM}-W_{ELM})}$. As discussed in section 8.2.4, the sampling distribution of r is not normal, therefore r is transformed to the quantity z in (8.5). Standard multilinear regression using least squares is then performed for yielding the regression coefficients given in table 8.6.

The regression model for CW plasmas is constructed using B_t, I_p,P_{input},n_e, Γ_{D_2} and δ_{avg} as predictor variables. For ILW plasmas, however, f_{STE} is included as an additional predictor variable, as it has been shown in section 8.3.1 that f_{STE} has an appreciable influence on $r_{(\Delta t_{ELM}-W_{ELM})}$. In addition, since f_{STE} is not strictly an engineering quantity, a second model (model 2) for ILW plasmas is constructed using Γ_{N_2} as an additional parameter in place of f_{STE}. The quality of the fitted regression model is quantified with the root-mean-square error (RMSE(%)), which is an indicator of the deviation of the measurements from the model, and the coefficient of determination ($R^2 \in [0,1]$), which measures the degree to which the predictor variables and the regression model explain the observed variation of the response variable. Based on the values of RMSE and R^2, each model is fairly appropriate to describe the variation of the correlation.

Across both model 1 and model 2 that are constructed for ILW plasmas, f_{STE} or alternatively Γ_{N_2} appear to be the most important determinant of $r_{(\Delta t_{ELM}-W_{ELM})}$.

Table 8.7: *Number of ILW plasmas (including N_2-seeded plasmas) and CW plasmas with correlation between energy loss of successive ELMs $r > 0.3$, $0.1 < r \leq 0.3$ and $-0.3 < r \leq 0.1$. The number of plasmas with r significantly different from zero are also indicated at two significance levels α.*

Plasmas	$-0.3 < r \leq 0.1$	$0.1 < r \leq 0.3$	$r > 0.3$	$r \neq 0$ ($\alpha = 5\%$)	$r \neq 0$ ($\alpha = 1\%$)
ILW	20	15	3	4	2
CW	16	4	0	3	0

This is expected since it has earlier been noted in section 8.3.1 that it is only with N_2 seeding that high values of $r_{(\Delta t_{ELM} - W_{ELM})}$ comparable with CW plasmas are obtained. In unseeded ILW plasmas the correlation fluctuates at most to a weakly positive correlation from a state of no correlation. Secondary to f_{STE}/Γ_{N_2}, δ_{avg} and Γ_{D_2} are the more important determinants of $r_{(\Delta t_{ELM} - W_{ELM})}$. This is consistent with the model for CW plasmas as therein δ_{avg} followed by Γ_{D_2} appear as the most important of the considered plasma engineering parameters. It is important to note that in addition to the global time-averaged plasma engineering parameters, the regression models could substantially benefit if the complete distributions of the predictor parameters would be considered.

8.5 Relation between energy loss of successive ELMs

Finally, the relationship between energy losses of consecutive ELMs is investigated. As can be noted from table 8.7, only 10 - 15 percent of the analyzed JET-ILW (including N_2-seeded plasmas) and JET-CW plasmas exhibit a weak non-zero correlation. Also, the values of r_s are in agreement with estimates of r. W_{ELM} of consecutive ELMs is largely uncorrelated. This implies that an ELM with a large W_{ELM} is equally likely to be followed by an ELM with a large or small W_{ELM}. Further, this observation is consistent across unseeded JET-ILW plasmas, N_2-seeded JET-ILW plasmas and JET-CW plasmas. This can also be observed in the scatter plots of W_{ELM} of nth ELM and W_{ELM} of $(n+1)th$ ELM in figure 8.13. For each of the three representative plasmas, #82806, #83179 and #76479, W_{ELM} of successive ELMs is uncorrelated.

8.6 Conclusions

This work examines the relation between W_{ELM} and Δt_{ELM} for individual ELMs in a set of non-seeded JET-ILW plasmas and compares the results with a set of N_2-seeded JET-ILW plasmas and JET-CW plasmas. It is found that the empirically established inverse relation between average f_{ELM} and $\bar{W}_{ELM}$ is not ubiquitously obeyed by individual ELMs. The linear correlation between W_{ELM} and Δt_{ELM} varies from being strongly correlated for certain plasmas to being completely uncorrelated for others. CW plasmas, in general, exhibit higher correlation between W_{ELM} and Δt_{ELM} than ILW plasmas and it is only in N_2-seeded ILW plasmas that a high correlation comparable to certain CW plasmas is observed.

Furthermore, ELMs in non-seeded JET ILW plasmas are often followed by a slow transport event resulting in a bi-modal distribution of ELM durations. The two modes correspond to two distinct underlying phenomena: pure ELMs and ELMs followed by a slow transport event. Slow transport events are not present in JET-CW plasmas and they disappear in N_2-seeded JET-ILW plasmas, giving rise to a unimodal asymmetric distribution of ELM durations. The average ELM energy loss in a plasma scales linearly with the proportion of ELMs followed by slow transport events in a plasma, whereas the linear correlation between W_{ELM} and Δt_{ELM} varies inversely with the fraction of slow transport events.

A collective analysis of all the ELMs from the unseeded JET-ILW ELMs plasmas revealed that the variation between W_{ELM} and Δt_{ELM} obeys a power law relationship. W_{ELM} appears to saturate for $\Delta t_{ELM} \approx 25 - 30ms$ which is roughly the time taken for the plasma thermal energy to return to its pre-ELM value. This suggests a scenario where the linear correlation between W_{ELM} and Δt_{ELM} significantly reduces as the edge pedestal recovers to its pre-ELM value.

Moreover, least squares linear regression has been employed for determining the region of the plasma operating regime where the correlation between W_{ELM} and Δt_{ELM} is maximized. A regression model is constructed using plasma and engineering parameters for both JET-ILW and JET-CW plasmas. While the models will certainly benefit from more informative predictors, they nevertheless indicate the more important parameters from the plasma parameters used as predictors. For the JET-ILW plasmas, Γ_{N_2} followed by δ_{avg} and Γ_{D_2} contribute most to the

correlation between W_{ELM} and Δt_{ELM}. Similarly, for JET-CW plasmas δ_{avg} and Γ_{D_2} appear to be the most important determinants of correlation.

Lastly it is acknowledged that W_{ELM} and Δt_{ELM} are stochastic quantities and a precise analysis of these quantities needs to effectively incorporate the uncertainty on these quantities. It has also been shown that the standard deviation of W_{ELM} and Δt_{ELM} increases linearly with the mean value. Analyzing W_{ELM} and Δt_{ELM} for individual ELMs subtly allows for the standard deviation in W_{ELM} and Δt_{ELM} to be accommodated and indeed reveals additional information. It is emphasized that analyzing complete probability distributions of W_{ELM}, Δt_{ELM}, τ_{ELM} and other plasma parameters will yield a more comprehensive picture and will thus form the basis of future investigations.

Chapter 9

Conclusions and outlook

In this thesis, we have developed pattern recognition methods in non-Euclidean spaces of probability distributions and applied them with a view on contributing to the systematic analysis of edge-localized modes (ELMs). The first part of this work concerns methodology development for rendering pattern recognition (advanced data analysis) methods which are apt for handling the challenges posed by nuclear fusion plasmas i.e. substantial measurement uncertainties and non-deterministic phenomena. In the second part, the analysis pins on ELMs, whose control and physics understanding is highly significant for the next step fusion devices. In this work, it has been consistently demonstrated that treating complete distributions of plasma parameters in contrast to average values is more informative and comprehensive.

We will now present the general conclusions and a summary of the contributions that have been made. Afterwards, we give an outlook towards possible improvements to and continuation of the current work.

9.1 Conclusions

9.1.1 Pattern recognition methods in spaces of probability distributions

It is recognized that in fusion plasmas, physical quantities are characterized by substantial measurement uncertainty and stochasticity. In the presence of uncertainty, probability theory provides a natural description of the raw data.

We start from the view point that the fundamental object resulting from the measurement process is a probability distribution, with every single measurement providing a sample from this distribution. Since patterns are in fact geometric constructs (clusters, discriminant surfaces), we need geometric concepts, in particular distance between probability distribution functions (PDFs), to carry out pattern recognition on probabilistic manifolds. To do this, we have employed the mathematical framework of information geometry which, based on the Fisher information, enables computation of a geodesic distance (GD) between PDFs.

Following this, we formulate several pattern classification and dimensionality reduction (visualization) methods in spaces of probability distributions (probabilistic manifolds). This includes extrapolation to the manifold setting of the k-nearest neighbour (kNN) and conformal predictor (CP) classifiers and multidimensional scaling (MDS) and landmark multidimensional scaling (LMDS) data visualization methods. Further two new classification schemes namely distance-to-centroid (D2C) and principal geodesic classifier (PGC) are developed.

The latter two methods have not yet been applied to fusion data. Nevertheless, they have been validated by their application to the classification and retrieval of colored texture images represented in the wavelet domain. It is shown that texture classification benefits significantly by effective utilization of the information residing in the rich spectral band correlation structure by joint modeling through multivariate distributions. Both D2C and PGC yield high classification accuracy at low computational complexity and benchmark GD as a well-suited distance measure between probability distributions.

9.1.2 Visualization of tokamak operational space

GD-based MDS results in an information visualization tool which can be used for representing the complete distributions of the multidimensional data characterizing the operational space of fusion devices onto two-dimensional (2D) maps. It is applied here for mapping the operational space of ELMs from carbon-wall (CW) JET plasmas onto 2D maps. The maps enable tracking of trends in plasma parameters across the operational space. In addition, they can also be used with reasonable accuracy for the prediction of the plasma parameters, including ELM types, at a certain location in the operational space. Furthermore, a computationally efficient

version of GD-based MDS is presented in the form of GD-based LMDS. It is shown that GD-based LMDS can be used to generate 2D or 3D maps at a very fast rate albeit with a slight compromise of accuracy.

9.1.3 Classification of ELM types

As outlined in chapter 4, various types of ELMs have been identified on an empirical and phenomenological basis. Herein, automated classification schemes for ELM types are developed with the aim of reducing the effort of ELM experts in identifying ELM types by providing fast and standardized ELM classification. To this end, a number of classification methodologies have been explored and put to use.

Firstly, a GD-based CP classifier is presented and applied for the classification of type I and small ELMs from the (ITPA) Global H-mode Confinement Database. The strength of the CP classifier lies in its ability to start with an almost empty training set. This minimizes the requirement of a training data set with correctly labeled ELM types and can be a specially useful property in the early phases of ITER operation. Further, CPs provide an estimate of the reliability and accuracy of their prediction and also identify the cases for which a decision could not be made reliably by categorizing them as 'ambiguous'. This offers an advantage over a spurious classification. Finally, despite the ITPA database providing limited information on the underlying probability distribution of the plasma parameters, it is well illustrated that a GD classification based on complete distribution of parameters is much more informative and correct than the classification based on mean parameter values or the Euclidean distance.

Next, discriminant analysis (DA) is used for parametric classification of type I and type III ELMs from a set of CW JET plasmas. Linear separation hyperplanes between type I and III ELMs are derived in terms of a set of global plasma parameters. This provides a simple predictive criterion related to physical knowledge which can be used for prediction as well as the study of ELM occurrence boundaries and ELM physics.

Since DA makes an assumption about the underlying class distribution and presently cannot be applied on the probabilistic manifold, hence, k-nearest neighbour (kNN) classifier is considered next. It is shown that kNN is a simple yet powerful classifier of ELM types. For the set of CW JET plasmas considered in this work

and using the global plasma parameters as predictors, GD-based kNN yields a classification accuracy which is comparable with DA. However, kNN is likely to outperform DA when the underlying class distribution is non-Gaussian. In the next step, a robust ELM detection algorithm, also developed in this doctoral work, is used for the extraction of inter-ELM time intervals (or waiting times) (Δt_{ELM}). kNN is employed for the classification of ELM types using distributions of Δt_{ELM}. It is clearly shown, that the distributions of Δt_{ELM} encompass more information than a single average value of ELM frequency (inverse mean ELM waiting time). It naturally follows that GD-based kNN using distributions of Δt_{ELM}, yields superior classification performance in comparison with mean Δt_{ELM} and the Euclidean distance measure.

9.1.4 Correlation analysis for ELM energy loss and waiting time

ELM control methods, in particular ELM pacing techniques, rely on the empirically observed inverse dependence of average ELM energy loss (averaged over a discharge) on average ELM frequency. However, the aim of ELM control is to reduce the size of individual ELMs and not the average loss. In this work, the correlation between ELM energy loss (W_{ELM}) and Δt_{ELM} is studied for individual ELMs in a set of ITER-like wall (ILW) JET plasmas. It is noted that while the average W_{ELM} and Δt_{ELM} conform to the empirically observed inverse dependence, for individual ELMs the correlation varies from zero to a moderately positive value. CW JET plasmas, in general, exhibit higher correlation between W_{ELM} and Δt_{ELM} than the ILW JET plasmas and it is seen that a correlation as high as that seen in CW plasmas is only observed in N_2-seeded ILW JET plasmas. It is shown that W_{ELM} and Δt_{ELM} exhibit a linear correlation until the pedestal has recovered to its pre-ELM value, beyond which W_{ELM} and Δt_{ELM} apper to be roughly independent of each other.

Most unseeded JET ILW plasmas have ELMs that are followed by a second collapse phase referred to as the slow transport event (STE). The influence of STEs on the distribution of ELM durations (τ_{ELM}) is studied and it is seen that the correlation between W_{ELM} and Δt_{ELM} has a weakly inverse relation with the fraction of slow transport events (f_{STS}).

Lastly, a regression analysis is performed for determining a region in the space of global plasma parameters where the correlation between W_{ELM} and Δt_{ELM} is maximized.

This analysis aims at presenting insights for an improved physics understanding of ELMs and an optimization of ELM control methods. Further, the value of studying the variation of the complete distribution of Δt_{ELM} with the distribution of W_{ELM} is highlighted and stressed.

9.2 Outlook

The work described in this thesis highlights the potential of advanced data analysis techniques, in particular pattern recognition methods, for enhancing the physics understanding and improving engineering and control of fusion plasmas. We here present an overview of the major findings in this work that can be improved, or which suggest further directions of the research.

The assembled data sets from JET and AUG, as well as the two international databases (ITPA database and VisTex texture database) treated in this thesis provide a reasonable basis for the analysis conducted in the work. However, an expansion of the current datasets with well diagnosed parameters and representative plasma discharges can further improve the robustness of analysis. A natural extension would then be an application (possibly in real-time) of the methods and analysis presented in this work for physics studies and plasma control at the current fusion devices.

It is important to emphasize that the methods developed in this work are generic and can be applied with relative ease to other phenomena in fusion plasmas such as plasma disruptions or turbulence studies. Further, as illustrated with the application of D2C and PGC classifiers to color texture discrimination, the methods can also be exported to and be useful in fields other than nuclear fusion.

Application of DA for ELM classification provides an exploratory step in developing recognition algorithms related to physical knowledge and well established empirical behaviour. However, DA at present cannot be applied in the spaces of probability distributions. This bars the use of DA classification from incorporating uncertainties in an optimal manner. Development of DA and other parametric

methods such as logistic regression, on the probabilistic manifolds is likely to be of value in a variety of domains including nuclear fusion.

Many extensions can also be suggested to the classification schemes presented and applied for ELM classification. The feature space can be expanded to include distributions of other plasma parameters and/or ELM properties, such as ELM energy losses and durations. However, in order to avoid the peaking phenomenon described in chapter 2, this must be accompanied with an increase in the number of analyzed plasma discharges. Secondly, the current classification schemes are perfectly capable in the current form of providing a machine independent classification of ELM types. However, this requires that the feature set be adjusted to include dimensionless plasma parameters or that the parameters be normalized with respect to the machine size or volume. Thirdly, the scope of the classification schemes can be expanded and they can be applied for the classification of other ELM types.

Lastly, in the light of the observation that the correlation between W_{ELM} and Δt_{ELM} fluctuates from zero to a moderately high value, a re-visit of the empirically observed relationship between average W_{ELM} and Δt_{ELM} using complete distributions of W_{ELM} and Δt_{ELM} can be fruitful.

Appendices

Appendix A

Databases

A.1 JET ELM-DBI

The given database, JET ELM-DBI, has been used in chapter 7 for the classification of ELM types.

In the following tables of data, the first column is the JET pulse number, t_1 and t_2 give the time at which the time series analysis of the relevant quantities started and ended respectively. Other parameters are as defined in chapter 7.

Table A.1: *JET ELM-DBI database: shot numbers, time interval for analysis and values of global plasma parameters.*

	Shot	t_1 (s)	t_2 (s)	B_t (T)	I_p (MA)	P_{input} (MW)	n_e $(10^{19}m^{-2})$	Γ_{D_2} $(10^{22}s^{-1})$	δ_{avg}
Type I ELMs									
1.	50564	62.0	67.0	1.91	1.87	8.21	3.20	0.00	0.211
2.	52149	59.0	62.0	2.68	2.50	13.9	8.46	0.87	0.424
3.	52508	59.5	63.0	2.63	2.42	17.3	5.70	0.00	0.437
4.	52511	59.8	62.8	2.63	2.41	17.0	6.95	3.98	0.426
5.	52513	59.5	62.8	2.63	2.41	17.1	6.53	4.53	0.419
6.	52516	59.8	62.8	2.43	2.31	15.6	6.59	4.41	0.424
7.	52517	59.8	62.8	2.43	2.31	15.9	7.85	4.34	0.424
8.	52518	59.8	62.8	2.43	2.31	13.0	8.16	4.33	0.416
9.	52519	60.7	63.7	2.43	2.31	16.0	7.84	4.45	0.416

Continued on next page

Continued from previous page

	Shot	t_1 (s)	t_2 (s)	B_t (T)	I_p (MA)	P_{input} (MW)	n_e $(10^{19}m^{-2})$	Γ_{D_2} $(10^{22}s^{-1})$	δ_{avg}
10.	52521	60.7	63.7	2.43	2.31	16.0	7.83	4.34	0.429
11.	53142	59.0	63.8	2.42	2.30	15.0	7.54	4.27	0.411
12.	56128	59.0	62.5	2.67	2.49	12.7	9.32	2.70	0.445
13.	56143	59.0	62.0	2.67	2.50	12.7	7.64	2.82	0.423
14.	56144	59.5	63.3	2.67	2.50	13.3	7.43	2.85	0.425
15.	56739	62.5	67.0	1.41	1.41	7.57	3.79	1.36	0.255
16.	56740	63.5	67.0	1.41	1.41	11.0	3.71	1.18	0.270
17.	57861	59.0	63.3	2.68	2.50	14.9	6.43	2.85	0.411
18.	57863	59.0	63.3	2.68	2.50	14.2	9.11	2.82	0.459
19.	57865	59.0	63.3	2.68	2.50	15.9	9.32	2.87	0.463
20.	57866	59.0	63.3	2.68	2.50	15.5	8.86	2.82	0.460
21.	57870	59.0	63.3	2.68	2.50	15.3	8.20	2.21	0.456
22.	57871	59.0	63.3	2.68	2.50	15.6	7.72	1.19	0.456
23.	57872	59.0	63.3	2.68	2.50	13.9	9.11	0.74	0.461
24.	57877	59.8	62.8	2.68	2.50	13.2	7.95	3.57	0.417
25.	57885	59.0	62.8	2.68	2.50	15.3	7.68	4.20	0.408
26.	57886	59.0	62.5	2.68	2.50	15.1	7.94	3.59	0.417
27.	57888	59.0	62.8	2.68	2.50	16.5	8.10	4.69	0.415
28.	57896	59.5	63.0	2.68	2.50	16.9	6.16	0.69	0.418
29.	59354	60.0	63.5	2.68	2.50	16.5	6.34	0.69	0.416
30.	60584	54.5	58.4	2.15	2.75	19.2	6.58	3.09	0.234
31.	60709	60.0	63.8	2.67	2.50	14.7	7.41	4.62	0.412
32.	61471	59.0	63.5	2.68	2.51	17.1	5.93	0.80	0.417
33.	61472	59.0	63.5	2.68	2.51	17.3	5.85	0.80	0.419
34.	61478	56.7	59.7	2.50	2.99	16.3	9.51	8.91	0.218
35.	61479	59.5	63.5	2.68	2.51	17.5	5.84	0.80	0.421
36.	61480	60.0	63.0	2.68	2.50	17.6	8.52	7.65	0.418

Continued on next page

Continued from previous page

	Shot	t_1 (s)	t_2 (s)	B_t (T)	I_p (MA)	P_{input} (MW)	n_e $(10^{19}m^{-2})$	Γ_{D_2} $(10^{22}s^{-1})$	δ_{avg}
37.	62216	60.0	63.0	2.42	2.01	12.6	4.60	1.08	0.261
38.	62220	57.0	61.0	3.04	3.00	19.8	6.18	0.56	0.260
39.	62221	57.0	61.0	3.04	3.00	19.8	6.56	2.48	0.257
40.	62222	57.5	60.5	3.04	3.00	19.9	7.17	5.19	0.252
41.	62224	57.5	61.0	3.04	3.00	18.5	7.53	6.46	0.250
42.	66111	58.0	63.0	2.67	2.48	16.7	6.97	0.82	0.488
43.	66115	58.0	63.0	2.67	2.48	16.9	8.51	0.82	0.488
44.	66116	59.0	63.0	2.67	2.48	16.7	7.88	0.37	0.484
45.	67761	59.5	63.0	2.67	2.49	16.8	6.54	0.00	0.447
46.	69373	63.5	66.5	1.69	1.99	15.8	5.01	0.67	0.418
47.	69900	55.5	59.3	2.79	2.99	22.1	9.86	7.54	0.396
48.	70050	56.0	59.7	2.91	2.98	20.8	8.80	4.24	0.405
49.	72339	59.0	63.0	2.67	2.48	13.9	5.75	1.19	0.400
50.	72343	58.5	63.3	2.67	2.48	14.6	6.51	2.70	0.405
51.	72345	60.0	63.0	2.66	2.48	14.4	5.89	0.00	0.434
52.	73087	59.5	63.3	2.66	2.48	15.2	6.26	2.64	0.396
53.	73335	59.0	63.0	2.66	2.48	14.7	6.34	0.00	0.433
54.	73341	59.0	63.0	2.66	2.48	15.7	5.96	1.50	0.400
55.	73345	59.5	63.0	2.66	2.48	14.2	5.31	1.16	0.402
56.	73346	59.0	63.0	2.66	2.48	15.0	5.50	2.20	0.400
57.	75722	65.0	69.5	1.58	1.51	8.18	3.18	0.08	0.291
58.	75727	64.0	69.0	1.99	1.97	12.8	4.39	0.00	0.275
59.	75731	64.5	67.5	1.99	1.98	12.9	3.88	0.00	0.278
60.	75732	64.5	67.5	1.99	1.98	12.9	3.69	0.00	0.285
61.	76473	58.5	61.5	1.99	1.98	15.4	4.88	0.08	0.280
62.	76474	58.0	61.5	1.99	1.98	15.6	4.00	0.07	0.275
63.	76475	58.5	61.5	1.99	1.98	15.6	3.74	0.07	0.272

Continued on next page

Continued from previous page

	Shot	t_1 (s)	t_2 (s)	B_t (T)	I_p (MA)	P_{input} (MW)	n_e $(10^{19}m^{-2})$	Γ_{D_2} $(10^{22}s^{-1})$	δ_{avg}
64.	76476	58.5	61.5	1.99	1.98	14.7	5.19	0.08	0.279
65.	76478	58.5	61.5	1.99	1.98	15.6	4.99	0.08	0.279
66.	76479	58.0	62.0	1.99	1.98	15.8	5.19	0.26	0.278
67.	76480	58.0	61.3	1.99	1.98	19.7	5.28	0.08	0.280
68.	76481	58.0	61.5	1.99	1.98	11.3	4.69	0.08	0.276
69.	76483	58.0	61.5	1.99	1.98	20.1	5.37	0.00	0.281
Type I high frequency (HF) ELMs									
70.	66109	59.0	63.0	2.67	2.49	18.7	5.77	0.00	0.495
71.	66108	59.0	62.5	2.67	2.49	18.2	5.36	0.00	0.486
72.	66107	59.0	63.0	2.67	2.49	19.2	6.06	0.20	0.491
73.	66106	59.0	63.0	2.67	2.49	18.0	5.95	0.00	0.482
74.	66105	59.0	63.0	2.67	2.49	18.0	6.12	0.00	0.488
Type III ELMs									
75.	68608	62.0	67.0	2.39	1.97	5.38	4.33	0.39	0.428
76.	68610	60.5	66.0	2.39	1.97	5.14	4.63	0.57	0.426
77.	68612	60.5	66.0	2.39	1.98	5.68	4.82	0.56	0.430
78.	68613	60.5	66.0	2.39	1.98	5.75	4.81	0.56	0.431
79.	68614	60.5	66.0	2.39	1.98	5.72	4.79	0.57	0.430
80.	68615	60.5	66.0	2.39	1.98	5.99	4.89	0.55	0.432
81.	68618	60.5	66.0	2.39	1.98	5.92	4.91	0.55	0.432
82.	68619	60.5	66.0	2.39	1.98	5.87	4.92	0.57	0.431
83.	74410	56.0	60.5	2.03	2.46	13.2	8.20	6.71	0.378
84.	74411	56.0	60.5	2.03	2.46	15.0	8.91	6.71	0.386
85.	74412	56.0	60.5	2.03	2.46	14.0	7.95	6.75	0.382
86.	74415	56.0	60.5	2.03	2.47	15.0	8.90	5.66	0.386
87.	74417	57.0	60.5	2.03	2.46	13.8	7.70	3.57	0.385
88.	74427	56.0	60.5	2.03	2.46	15.0	7.99	6.74	0.389

Continued on next page

Continued from previous page

	Shot	t_1 (s)	t_2 (s)	B_t (T)	I_p (MA)	P_{input} (MW)	n_e $(10^{19}m^{-2})$	Γ_{D_2} $(10^{22}s^{-1})$	δ_{avg}
89.	74428	57.0	60.5	2.03	2.46	15.0	7.31	6.31	0.388
90.	70136	61.0	63.0	1.74	2.28	5.11	1.65	0.09	0.214
91.	68748	48.3	51.3	1.69	1.69	21.9	4.89	5.49	0.416
92.	68743	48.7	51.1	1.69	1.69	17.7	5.57	8.43	0.417
93.	74661	54.5	57.5	2.37	2.61	8.56	2.78	0.00	0.253
94.	50567	62.0	66.0	1.91	1.86	7.73	6.00	6.38	0.197
95.	56131	58.5	61.5	2.67	2.48	14.1	9.80	8.08	0.440
96.	73080	61.0	63.0	2.66	2.47	18.9	5.44	0.87	0.418
97.	74519	57.1	59.1	2.63	3.24	19.4	10.50	6.90	0.394
98.	74429	57.5	59.0	2.43	2.96	17.9	9.56	8.16	0.390
99.	74431	57.6	58.6	2.43	2.97	18.2	9.81	8.23	0.393
100.	74513	57.0	59.0	2.43	2.99	17.8	9.75	8.02	0.388

Table A.2: *JET ELM-DBI database: shot numbers, time interval for analysis and parameter estimates for Gaussian distribution and 2P-Weibull distribution fits to the ELM waiting times (Δt_{ELM}).*

	Shot	t_1 (s)	t_2 (s)	Δt_{ELM}			
				Gaussian		2P Weibull	
				μ $(10^{-2}\ s)$	σ $(10^{-2}\ s)$	α $(10^{-2}\ s)$	β
Type I ELMs							
1	50564	62.0	67.0	4.39	0.805	4.72	5.67
2	52149	59.0	62.0	4.16	0.459	4.36	9.78
3	52508	59.5	63.0	5.77	3.15	6.50	1.89
4	52511	59.8	62.8	2.68	0.615	2.90	5.45
5	52513	59.5	62.8	2.16	0.477	2.35	5.08
6	52516	59.8	62.8	2.12	0.599	2.34	3.67
7	52517	59.8	62.8	4.09	1.10	4.51	4.07

Continued on next page

Continued from previous page

	Shot	t_1 (s)	t_2 (s)	Δt_{ELM}			
				Gaussian		2P Weibull	
				μ $(10^{-2}\ s)$	σ $(10^{-2}\ s)$	α	β
8	52518	59.8	62.8	6.46	1.40	7.02	5.08
9	52519	60.7	63.7	5.02	1.22	5.50	4.28
10	52521	60.7	63.7	4.68	1.28	5.13	3.97
11	53142	59.0	63.8	2.35	0.676	2.60	3.54
12	56128	59.0	62.5	3.56	0.543	3.80	6.64
13	56143	59.0	62.0	4.06	0.751	4.38	5.36
14	56144	59.5	63.3	3.29	0.696	3.58	4.57
15	56739	62.5	67.0	4.35	0.735	4.67	6.07
16	56740	63.5	67.0	1.66	0.322	1.79	4.79
17	57861	59.0	63.3	2.66	0.373	2.81	8.38
18	57863	59.0	63.3	4.10	0.925	4.47	4.37
19	57865	59.0	63.3	4.15	1.17	4.58	3.70
20	57866	59.0	63.3	3.16	0.887	3.50	3.71
21	57870	59.0	63.3	2.62	0.669	2.87	4.62
22	57871	59.0	63.3	2.76	0.767	3.02	4.47
23	57872	59.0	63.3	6.18	0.798	6.55	7.31
24	57877	59.8	62.8	5.27	0.918	5.64	6.56
25	57885	59.0	62.8	3.99	1.18	4.41	3.63
26	57886	59.0	62.5	4.55	0.687	4.86	6.20
27	57888	59.0	62.8	3.80	1.32	4.23	2.69
28	57896	59.5	63.0	3.13	0.832	3.44	3.57
29	59354	60.0	63.5	3.87	1.22	4.30	3.10
30	60584	54.5	58.4	5.24	1.40	5.77	3.94
31	60709	60.0	63.8	3.02	0.590	3.26	5.79
32	61471	59.0	63.5	2.07	0.346	2.22	6.45

Continued on next page

Continued from previous page

	Shot	t_1 (s)	t_2 (s)	Δt_{ELM}			
				Gaussian		2P Weibull	
				μ $(10^{-2}\ s)$	σ $(10^{-2}\ s)$	α	β
33	61472	59.0	63.5	2.14	0.405	2.31	5.65
34	61478	56.7	59.7	2.34	1.08	2.64	2.28
35	61479	59.5	63.5	2.18	0.437	2.35	6.02
36	61480	60.0	63.0	2.79	1.78	3.09	1.53
37	62216	60.0	63.0	3.03	1.14	3.41	2.92
38	62220	57.0	61.0	5.89	1.47	6.44	3.95
39	62221	57.0	61.0	4.27	0.840	4.62	5.11
40	62222	57.5	60.5	3.02	1.70	3.39	1.82
41	62224	57.5	61.0	2.74	1.55	3.05	1.73
42	66111	58.0	63.0	2.86	0.716	3.07	4.81
43	66115	58.0	63.0	2.67	0.872	2.91	3.39
44	66116	59.0	63.0	1.95	0.245	2.05	9.51
45	67761	59.5	63.0	1.34	0.131	1.40	11.6
46	69373	63.5	66.5	3.51	0.886	3.86	4.04
47	69900	55.5	59.3	3.07	2.31	3.32	1.30
48	70050	56.0	59.7	2.97	1.27	3.35	2.55
49	72339	59.0	63.0	3.65	0.461	3.86	8.66
50	72343	58.5	63.3	2.87	0.334	3.02	9.10
51	72345	60.0	63.0	2.27	0.423	2.44	6.27
52	73087	59.5	63.3	2.85	0.609	3.08	4.91
53	73335	59.0	63.0	2.50	0.590	2.73	4.31
54	73341	59.0	63.0	3.17	0.466	3.37	6.39
55	73345	59.5	63.0	3.14	0.549	3.35	6.95
56	73346	59.0	63.0	2.75	0.435	2.93	6.77
57	75722	65.0	69.5	1.77	0.735	1.99	2.42

Continued on next page

Continued from previous page

	Shot	t_1 (s)	t_2 (s)	Δt_{ELM}			
				Gaussian		2P Weibull	
				μ $(10^{-2}\ s)$	σ $(10^{-2}\ s)$	α	β
58	75727	64.0	69.0	5.78	1.29	6.24	5.30
59	75731	64.5	67.5	4.41	1.27	4.89	3.91
60	75732	64.5	67.5	3.91	1.24	4.35	3.43
61	76473	58.5	61.5	3.62	0.695	3.89	5.74
62	76474	58.0	61.5	3.67	0.755	3.95	6.08
63	76475	58.5	61.5	3.62	1.25	4.03	3.41
64	76476	58.5	61.5	3.66	1.53	4.11	2.64
65	76478	58.5	61.5	3.48	1.12	3.86	3.66
66	76479	58.0	62.0	2.86	1.24	3.20	2.44
67	76480	58.0	61.3	2.51	1.13	2.82	2.37
68	76481	58.0	61.5	4.57	1.25	5.01	4.09
69	76483	58.0	61.5	2.45	1.28	2.74	1.94
Type I *high frequency* (HF) ELMs							
70	66109	59.0	63.0	0.762	0.225	0.844	3.39
71	66108	59.0	62.5	0.753	0.257	0.839	2.89
72	66107	59.0	63.0	0.755	0.165	0.822	4.08
73	66106	59.0	63.0	1.10	0.300	1.21	3.58
74	66105	59.0	63.0	0.759	0.181	0.828	3.46
Type III ELMs							
75	68608	62.0	67.0	0.371	0.402	0.379	1.05
76	68610	60.5	66.0	0.447	0.418	0.482	1.23
77	68612	60.5	66.0	0.638	0.682	0.676	1.15
78	68613	60.5	66.0	0.606	0.487	0.665	1.35
79	68614	60.5	66.0	0.673	0.597	0.745	1.37
80	68615	60.5	66.0	0.660	0.537	0.730	1.39

Continued on next page

Continued from previous page

	Shot	t_1 (s)	t_2 (s)	Δt_{ELM}			
				Gaussian		2P Weibull	
				μ $(10^{-2}\ s)$	σ $(10^{-2}\ s)$	α	β
81	68618	60.5	66.0	0.666	0.533	0.740	1.42
82	68619	60.5	66.0	0.653	0.650	0.717	1.31
83	74410	56.0	60.5	0.371	0.218	0.419	1.81
84	74411	56.0	60.5	0.279	0.283	0.309	1.35
85	74412	56.0	60.5	0.213	0.287	0.211	0.98
86	74415	56.0	60.5	0.196	0.121	0.220	1.68
87	74417	57.0	60.5	0.189	0.192	0.203	1.23
88	74427	56.0	60.5	0.218	0.160	0.244	1.51
89	74428	57.0	60.5	0.214	0.181	0.232	1.27
90	70136	61.0	63.0	1.46	0.707	1.66	2.21
91	68748	48.3	51.3	0.695	0.587	0.761	1.32
92	68743	48.7	51.1	0.534	0.518	0.596	1.42
93	74661	54.5	57.5	1.23	0.749	1.36	1.58
94	50567	62.0	66.0	2.23	1.37	2.47	1.55
95	56131	58.5	61.5	0.361	0.181	0.408	2.11
96	73080	61.0	63.0	0.402	0.408	0.433	1.21
97	74519	57.1	59.1	0.776	1.00	0.800	1.06
98	74429	57.5	59.0	0.793	0.633	0.872	1.36
99	74431	57.6	58.6	0.197	0.158	0.216	1.35
100	74513	57.0	59.0	0.280	0.168	0.316	1.80

A.2 AUG ELM-DBI

Table A.3: *AUG ELM-DBI database: shot numbers, time interval for analysis and parameter estimates for Gaussian distribution and 2P-Weibull distribution fits to the ELM waiting times (Δt_{ELM})*

	Shot	t_1 (s)	t_2 (s)	Δt_{ELM}			
				Gaussian		2P Weibull	
				μ $(10^{-2}\ s)$	σ $(10^{-2}\ s)$	α $(10^{-2}\ s)$	β
Type I ELMs							
1	29209	2.00	4.00	2.29	0.537	2.50	4.49
2	29212	2.00	4.00	1.60	0.400	1.75	3.85
3	29197	2.10	3.60	1.79	0.661	2.01	2.97
4	29526	3.00	5.00	4.20	0.154	4.71	2.84
5	30410	5.10	5.80	1.49	0.491	1.66	3.48
6	30465	5.00	5.80	1.51	0.522	1.69	3.20
7	30479	5.00	5.80	1.36	0.636	1.54	2.30
8	30525	5.10	5.80	1.46	0.404	1.61	4.27
9	30564	2.20	4.00	0.809	0.351	0.91	2.47
10	30587	5.10	5.80	1.44	0.701	1.63	2.24
11	30658	2.00	2.80	1.24	0.208	1.33	5.75
12	30658	3.70	4.70	1.22	0.448	1.35	3.00
13	29191	2.00	4.00	1.54	0.534	1.68	2.83
14	29192	2.00	4.00	1.46	0.523	1.62	2.89
15	29208	2.00	4.00	1.65	0.413	1.80	4.25
16	29211	2.50	4.00	1.54	0.361	1.68	4.39
17	29904	2.50	3.50	0.765	0.367	0.87	2.19
18	29904	4.00	5.00	0.798	0.306	0.90	2.71
19	31447	3.50	5.00	0.893	0.317	1.00	3.16
20	31137	2.20	3.20	3.02	1.07	3.31	2.80

Continued on next page

Continued from previous page

	Shot	t_1 (s)	t_2 (s)	Δt_{ELM}			
				Gaussian		2P Weibull	
				μ $(10^{-2}\ s)$	σ $(10^{-2}\ s)$	α	β
Type III/Mixed ELMs							
21	30410	2.00	4.50	0.530	0.205	0.60	2.72
22	31170	2.00	3.50	0.436	0.0906	0.47	5.00
23	30465	2.00	4.00	0.592	0.257	0.67	2.39
24	30479	2.00	4.00	0.599	0.250	0.68	2.48
25	30521	2.00	4.00	0.499	0.181	0.56	2.88
26	30525	2.00	4.00	0.569	0.239	0.64	2.45
27	30587	3.00	5.00	0.665	0.437	0.75	1.68
28	30628	2.00	4.00	0.503	0.201	0.57	2.59
29	31447	2.00	3.00	0.737	0.449	0.83	1.74
30	31499	2.00	5.00	0.762	0.478	0.86	1.72

A.3 JET ELM-DBII

The given database, JET ELM-DBII, has been used in chapter 8 for the correlation analysis of energy losses, waiting times and durations for type I ELMs in JET.

In the following tables of data, the first column is the JET pulse number, t_1 and t_2 give the time at which the time series analysis of the relevant quantities started and ended respectively. Other parameters are as defined in chapter 8.

Table A.4: *JET ELM-DBII database: shot numbers, time interval for analysis and values of global plasma parameters.*

	Shot	t_1 (s)	t_2 (s)	B_t (T)	I_p (MA)	P_{input} (MW)	n_e $(10^{19}m^{-2})$	Γ_{D_2} $(10^{22}s^{-1})$	δ_{avg}
JET-ILW plasmas									
1.	83630	48.5	51.5	2.00	2.00	12.5	4.62	1.18	0.275
2.	83630	52.0	54.0	2.00	2.00	10.9	5.03	1.15	0.276
3.	83631	51.5	53.2	2.00	2.00	12.2	4.9	1.16	0.277
4.	83632	50.0	53.0	2.00	2.00	12.5	4.69	1.17	0.275
5.	83633	48.5	51.5	2.00	2.00	12.5	4.70	1.19	0.275
6.	83634	50.0	53.0	2.00	2.00	12.4	5.01	1.18	0.278
7.	83635	49.0	52.0	2.00	2.00	12.5	4.82	1.18	0.276
8.	83637	50.0	53.0	2.00	2.00	12.7	5.01	1.16	0.278
9.	83640	50.0	53.0	2.00	2.00	12.5	4.76	1.17	0.276
10.	83641	50.0	53.0	2.00	2.00	12.5	4.79	1.17	0.276
11.	83642	49.0	53.5	2.00	2.00	12.5	4.78	1.18	0.275
12.	83337	55.0	56.8	1.60	1.60	13.5	3.24	0.690	0.290
13.	82806	54.0	57.0	2.65	2.50	16.5	6.03	2.88	0.371
14.	82537	54.0	56.0	2.65	2.50	16.50	5.95	3.47	0.376
15.	82781	54.0	56.0	2.60	2.50	17.8	4.49	1.17	0.273
16.	83177	53.0	56.0	2.70	2.50	16.2	5.66	1.41	0.267
17.	82755	56.2	57.2	2.65	2.50	15.2	6.17	2.92	0.371
18.	82536	54.0	56.0	2.60	2.50	16.9	6.48	3.95	0.374
19.	82540	54.0	56.0	2.65	2.50	15.8	5.59	1.91	0.377

Continued on next page

Continued from previous page

	Shot	t_1 (s)	t_2 (s)	B_t (T)	I_p (MA)	P_{input} (MW)	n_e $(10^{19}m^{-2})$	Γ_{D_2} $(10^{22}s^{-1})$	δ_{avg}
20.	82541	54.0	56.0	2.65	2.50	16.7	5.24	1.55	0.381
21.	82546	54.0	56.5	2.60	2.50	15.4	6.26	3.01	0.374
22.	82550	54.0	57.0	2.60	2.50	15.1	6.07	1.92	0.374
23.	82900	58.3	60.3	2.41	1.30	8.36	1.97	0.532	0.287
24.	82900	60.7	62.7	2.41	1.30	13.5	1.87	0.520	0.300
25.	82898	60.5	62.5	2.41	1.60	13.3	2.51	0.627	0.294
26.	83548	54.0	55.5	1.56	1.40	13.3	2.89	0.905	0.412
27.	83537	54.8	56.0	2.36	1.50	15.3	2.55	1.02	0.399
28.	83339	55.5	57.0	1.30	1.30	14.5	2.40	0.564	0.299
29.	82549	54.0	56.0	2.65	2.50	15.4	6.20	3.12	0.374
30.	82550	54.0	57.0	2.65	2.50	15.1	6.07	1.92	0.374
31.	82636	57.7	59.7	1.70	1.60	8.43	3.71	0.669	0.284
32.	82635	57.9	59.9	1.30	1.30	6.92	2.97	0.569	0.287
JET-ILW plasmas, N_2 seeded									
33.	82812	54.0	57.0	2.65	2.50	19.1	7.01	2.69	0.391
34.	82813	54.0	57.0	2.65	2.50	16.9	6.52	1.27	0.388
35.	83179	53.0	55.0	2.70	2.50	16.7	5.43	1.36	0.272
36.	83180	53.0	55.0	2.70	2.50	17.5	5.36	1.26	0.271
37.	82811	54.0	57.0	2.65	2.50	17.0	6.46	3.66	0.382
38.	82810	54.0	57.0	2.65	2.50	17.0	7.39	2.66	0.388
JET-CW plasmas									
1.	76475	58.5	61.5	2.00	2.00	15.5	3.74	0.0745	0.272
2.	69900	55.5	59.3	2.80	3.00	21.7	9.86	7.54	0.396
3.	70050	56.0	59.7	2.90	3.00	20.7	8.80	4.24	0.405
4.	76478	58.5	61.5	2.00	2.00	15.3	4.99	0.0774	0.279
5.	72343	58.5	60.5	2.70	2.50	14.5	6.58	2.76	0.405
6.	72345	60.0	63.0	2.70	2.50	14.3	5.89	0.00	0.434

Continued on next page

Continued from previous page

	Shot	t_1 (s)	t_2 (s)	B_t (T)	I_p (MA)	P_{input} (MW)	n_e $(10^{19}m^{-2})$	Γ_{D_2} $(10^{22}s^{-1})$	δ_{avg}
7.	73335	59.0	63.0	2.70	2.50	14.5	6.34	0.00	0.433
8.	73341	59.0	63.0	2.70	2.50	15.5	5.96	1.50	0.400
9.	76481	58.0	61.5	2.00	2.00	11.0	4.69	0.0778	0.276
10.	75722	65.0	69.5	1.60	1.50	8.11	3.18	0.0820	0.291
11.	75727	64.0	69.0	2.00	2.00	12.6	4.39	0.00	0.275
12.	76473	58.5	61.5	2.00	2.00	15.2	4.88	0.0779	0.280
13.	76474	58.0	61.5	2.00	2.00	15.3	4.00	0.0750	0.275
14.	69373	63.5	66.5	1.70	2.00	15.5	5.01	0.675	0.418
15.	76476	58.5	61.5	2.00	2.00	14.3	5.19	0.0771	0.279
16.	72339	59.0	63.0	2.70	2.50	13.7	5.75	1.19	0.400
17.	76479	58.0	62.0	2.00	2.00	15.5	5.19	0.256	0.278
18.	76480	58.0	61.3	2.00	2.00	19.6	5.28	0.0785	0.280
19.	73345	59.5	63.0	2.70	2.50	14.1	5.31	1.16	0.400
20.	76483	58.0	61.5	2.00	2.00	19.7	5.37	0.00	0.281

Table A.5: *JET ELM-DBII database: shot numbers, time interval for analysis, mean and coefficient of variation for ELM waiting time, energy loss and duration for each plasma.*

	Shot	t_1	t_2	Δt_{ELM}		W_{ELM}		τ_{ELM}	
				μ $(10^{-2}\ s)$	σ/μ	μ $(10^{5}\ J)$	σ/μ	μ $(10^{-2}\ s)$	σ/μ
JET-ILW plasmas									
1	82537	48.5	51.5	4.46	0.510	2.30	0.682	0.595	0.777
2	83630	52.0	54.0	4.42	0.411	1.27	0.179	0.310	0.207
3	83631	51.5	53.2	3.31	0.374	1.21	0.179	0.307	0.290
4	83632	50.0	53.0	2.72	0.444	1.10	0.209	0.335	0.464
5	83633	48.5	51.5	3.00	0.421	1.18	0.236	0.311	0.269
6	83634	50.0	53.0	3.36	0.390	1.10	0.243	0.265	0.204
7	83635	49.0	52.0	3.22	0.341	1.23	0.168	0.283	0.218
8	83637	50.0	53.0	3.11	0.325	1.12	0.170	0.252	0.208
9	83640	50.0	53.0	2.88	0.399	1.11	0.209	0.293	0.244
10	83641	50.0	53.0	2.98	0.398	1.16	0.182	0.300	0.255
11	83642	49.0	53.5	3.10	0.403	1.16	0.204	0.298	0.246
12	83337	55.0	56.8	2.03	0.416	0.639	0.337	0.259	0.210
13	82806	54.0	57.0	5.53	0.317	3.12	0.455	0.826	0.513
14	83630	54.0	56.0	3.07	0.377	1.23	0.186	0.312	0.273
15	82781	54.0	56.0	3.43	0.387	2.47	0.316	0.520	0.399
16	83177	53.0	56.0	4.22	0.398	1.85	0.294	0.351	0.394
17	82755	56.2	57.2	4.37	0.319	2.02	0.618	0.548	0.85
18	82536	54.0	56.0	3.33	0.387	1.81	0.641	0.549	0.686
19	82540	54.0	56.0	6.23	0.349	3.13	0.388	0.770	0.506
20	82541	54.0	56.0	5.80	0.321	3.23	0.383	0.794	0.492
21	82546	54.0	56.5	4.08	0.396	2.05	0.594	0.608	0.662
22	82550	54.0	57.0	5.17	0.426	2.49	0.538	0.626	0.586
23	82900	58.3	60.3	1.62	0.432	0.320	0.544	0.321	0.316
24	82900	60.7	62.7	0.962	0.322	0.256	0.543	0.302	0.204

Continued on next page

Continued from previous page

	Shot	t_1	t_2	Δt_{ELM}		W_{ELM}		τ_{ELM}	
				μ $(10^{-2}\ s)$	σ/μ	μ $(10^{5}\ J)$	σ/μ	μ $(10^{-2}\ s)$	σ/μ
25	82898	60.5	62.5	2.12	0.312	0.679	0.511	0.278	0.267
26	83548	54.0	55.5	2.28	0.302	0.824	0.521	0.265	0.304
27	83537	54.8	56.0	1.08	0.337	0.378	0.505	0.257	0.151
28	83339	55.5	57.0	0.925	0.239	0.278	0.331	0.206	0.206
29	82549	54.0	56.0	4.45	0.335	2.09	0.551	0.487	0.619
30	82550	54.0	57.0	5.12	0.444	2.43	0.562	0.564	0.661
31	82636	57.7	59.7	4.21	0.216	0.734	0.244	0.222	0.185
32	82635	57.9	59.9	2.72	0.244	0.395	0.308	0.218	0.249
JET-ILW plasmas, N_2 seeded									
33	82812	54.0	57.0	4.87	0.364	1.64	0.236	0.264	0.217
34	82813	54.0	57.0	5.37	0.520	1.80	0.291	0.256	0.226
35	83179	53.0	55.0	3.58	0.400	1.25	0.310	0.281	0.207
36	83180	53.0	55.0	3.40	0.431	1.37	0.455	0.352	0.277
37	82811	54.0	57.0	2.48	0.639	0.760	0.422	0.196	0.262
38	82810	54.0	57.0	6.82	0.505	1.72	0.332	0.232	0.219
JET-CW plasmas									
1	76475	58.5	61.5	3.62	0.346	2.33	0.220	0.186	0.499
2	69900	55.5	59.3	3.65	0.658	2.01	0.505	0.381	0.382
3	70050	56.0	59.7	2.97	0.427	1.85	0.378	0.263	0.473
4	76478	58.5	61.5	3.48	0.321	2.73	0.166	0.222	0.224
5	72343	58.5	60.5	3.06	0.111	1.07	0.182	0.225	0.141
6	72345	60.0	63.0	2.27	0.186	1.05	0.345	0.339	0.607
7	73335	59.0	63.0	2.50	0.236	1.19	0.312	0.259	0.618
8	73341	59.0	63.0	3.17	0.147	1.31	0.243	0.289	0.234
9	76481	58.0	61.5	4.57	0.274	2.58	0.163	0.162	0.205
10	75722	65.0	69.5	1.77	0.416	0.395	0.329	0.252	0.175

Continued on next page

Continued from previous page

	Shot	t_1	t_2	Δt_{ELM}		W_{ELM}		τ_{ELM}	
				μ $(10^{-2}\ s)$	σ/μ	μ $(10^{5}\ J)$	σ/μ	μ $(10^{-2}\ s)$	σ/μ
11	75727	64.0	69.0	5.78	0.224	3.18	0.122	0.210	0.324
12	76473	58.5	61.5	3.62	0.192	2.81	0.102	0.245	0.199
13	76474	58.0	61.5	3.67	0.206	2.57	0.135	0.250	0.288
14	69373	63.5	66.5	3.51	0.252	1.89	0.295	0.269	0.293
15	76476	58.5	61.5	3.66	0.417	2.70	0.251	0.269	0.229
16	72339	59.0	63.0	3.65	0.126	1.62	0.263	0.196	0.192
17	76479	58.0	62.0	2.86	0.432	2.14	0.265	0.245	0.447
18	76480	58.0	61.3	2.51	0.449	2.56	0.256	0.315	0.443
19	73345	59.5	63.0	3.14	0.175	1.27	0.324	0.245	0.615
20	76483	58.0	61.5	2.65	0.508	2.49	0.259	0.203	0.303

JET acronym for experiment	Experiment	No. of pulses
JET ILW plasmas		
Ex 3.2.2	ELM Physics	7
Ex 1.2.5	Pulses prior to LTS retrieval	11
Ex 1.3.2	Fueling and seeding studies	13
TFE2 EX 3.2.2	ELM physics, energy and heat loads scalings	2
Ex 2.1.5	Baseline scenario	5
JET CW plasmas		
—-	Development of high delta configuration	3
—-	High current, high delta operations	2
S1 2.4.1	Configuration development for low and high delta	3
TFE 2.4.1	ELM power balance	2
E 2.4.1	Characterization of regular ELMs, first wall load	9

Table A.6: *List of experiments to which the analyzed shots belong to. The left column reports the acronym for JET experiments.*

Appendix B

Frequently used parameters

A_{ELM}	ELM energy deposition area
a	Plasma minor radius
B	Magnetic field, has both toroidal (B_t) and poloidal (B_p) components
β_p	Poloidal beta
Γ_{D_2}	Deuterium gas fueling rate
Γ_{N_2}	Nitrogen gas fueling rate
Δt_{ELM}	Inter-ELM time interval or waiting time
δ	Plasma triangularity
E	Electric field
f_{ELM}	ELM repetition frequency
I_{div}	Scrape-off-layer current measured via shunt resistances through the divertor
I_p	Plasma current
$\lambda_{e,e}$	Mean electron-electron Coulomb collision mean free path length
$\mu_\circ$	Vacuum permeability
n_e	Electron density
q_{95}	Value of safety factor at 95% of normalized flux (close to plasma edge)
$R_\circ$	Plasma major radius
τ_E	Energy confinement time i.e. ratio of stored energy to input power
τ_{ELM}	ELM duration
$T_{e,i}$	Electron/Ion temperature (keV)
ν^*	Normalized collision frequency, also referred to as "collisionality"

ν^*_{ped}	Pedestal collisionality
P_{ICRH}	Ion cyclotron resonance heating power
P_{NBI}	Neutral beam injected heating power
P_{ohmic}	Ohmic heating power
P_{sep}	Energy flux through the separatix
q_{max}	Peak ELM energy flux
V	Plasma volume
W_{plasma}	Plasma stored energy
W_{ELM}	ELM energy loss
W_{MHD}	Plasma stored energy (kinetics)
W_{ped}	Pedestal energy

Bibliography

[1] L. Clarke, J. Edmonds, V. Krey, R. Richels, S. Rose, and M. Tavoni, "International climate policy architectures: Overview of the EMF 22 International Scenarios," *Energy Economics*, vol. 31, no. SUPPL. 2, pp. 64–81, 2009.

[2] "United nations conference on climate change," http://www.cop21.gouv.fr/en/learn/what-is-cop21/, 2015.

[3] J. Wesson, *Tokamaks*, 4th ed. Oxford University Press, 2011.

[4] "ITER organisation," https://www.iter.org/.

[5] A. Einstein, "Ist die Trägheit eines Körpers von seinem Energieinhalt abhängig?" *Annalen der Physik*, vol. 18, no. 4, pp. 639–641, 1905.

[6] F.F.Chen, *An indispensible truth, How fusion power can save the planet.* Springer, 2011.

[7] H.-S. Bosch and G. Hale, "Improved formulas for fusion cross-sections and thermal reactivities," *Nuclear Fusion*, vol. 32, no. 4, pp. 611–631, 1992.

[8] V. Smirnov, "Tokamak foundation in USSR/Russia 1950–1990," *Nuclear Fusion*, vol. 50, no. 1, p. 014003, 2009.

[9] "EUROfusion," https://www.euro-fusion.org/.

[10] M. Merola, D. Loesser, A. Martin, P. Chappuis, R. Mitteau, V. Komarov, R. A. Pitts, S. Gicquel, V. Barabash, L. Giancarli, J. Palmer, M. Nakahira, A. Loarte, D. Campbell, R. Eaton, A. Kukushkin, M. Sugihara, F. Zhang, C. S. Kim, R. Raffray, L. Ferrand, D. Yao, S. Sadakov, A. Furmanek, V. Rozov, T. Hirai, F. Escourbiac, T. Jokinen, B. Calcagno, and S. Mori,

"ITER plasma-facing components," *Fusion Engineering and Design*, vol. 85, no. 10-12, pp. 2312–2322, 2010.

[11] R. A. Pitts, S. Carpentier, F. Escourbiac, T. Hirai, V. Komarov, A. S. Kukushkin, S. Lisgo, A. Loarte, M. Merola, R. Mitteau, A. R. Raffray, M. Shimada, and P. C. Stangeby, "Physics basis and design of the ITER plasma-facing components," *Journal of Nuclear Materials*, vol. 415, no. 1 SUPPL, pp. S957–S964, 2011.

[12] G. F. Matthews, J.-E. contributors, and ASDEX Upgrade Team, "Plasma operation with an all metal first-wall: Comparison of an ITER-like wall with a carbon wall in JET," *Journal of Nuclear Materials*, vol. 438, no. SUPPL, 2013.

[13] R. Neu, V. Bobkov, R. Dux, J. C. Fuchs, O. Gruber, A. Herrmann, A. Kallenbach, H. Maier, M. Mayer, T. Pütterich, V. Rohde, A. C. C. Sips, J. Stober, K. Sugiyama, and ASDEX Upgrade Team, "Ten years of W programme in ASDEX Upgrade—challenges and conclusions," *Physica Scripta*, vol. T138, no. T138, p. 014038, 2009.

[14] F. Wagner, G. Becker, K. Behringer, D. Campbell, A. Eberhagen, W. Engelhardt, G. Fussmann, O. Gehre, J. Gernhardt, G. V. Gierke, G. Haas, M. Huang, F. Karger, M. Keilhacker, O. Kluber, M. Kornherr, K. Lackner, G. Lisitano, G. G. Lister, H. M. Mayer, D. Meisel, E. R. Muller, H. Murmann, H. Niedermeyer, W. Poschenrieder, H. Rapp, H. Rohr, F. Schneider, G. Siller, E. Speth, A. Stabler, K. H. Steuer, G. Venus, O. Vollmer, and Z. Yu, "Regime of improved confinement and high beta in neutral-beam-heated divertor discharges of the ASDEX tokamak," *Physical Review Letters*, vol. 49, no. 19, pp. 1408–1412, 1982.

[15] F. Wagner, "A quarter-century of H-mode studies," *Plasma Physics and Controlled Fusion*, vol. 49, no. 12B, pp. B1–B33, 2007.

[16] J. Connor and H. Wilson, "A review of theories of the L-H transition," *Plasma Physics and Controlled Fusion*, vol. 42, no. 1, pp. R1–R74, 2000.

[17] J. Connor, T. Fukuda, X. Garbet, C. Gormezano, V. Mukhovatov, M. Wakatani, the ITB Database Group, and the ITPA Topical Group on Transport and Internal Barrier Physics, "A review of internal transport barrier physics for steady-state operation of tokamaks," *Nuclear Fusion*, vol. 44, no. 4, pp. R1–R49, 2004.

[18] V. V. Parail, "Energy and particle transport in plasmas with transport barriers," *Plasma Physics and Controlled Fusion*, vol. 44, no. 5A, 2002.

[19] J. W. Connor, "Edge-localized modes - physics and theory," *Plasma Physics and Controlled Fusion*, vol. 40, no. 5, pp. 531–542, 1999.

[20] H. Zohm, "Edge localized modes (ELMs)," *Plasma Physics and Controlled Fusion*, vol. 38, no. 2, pp. 105–128, 1996.

[21] J. W. Coenen, G. Arnoux, B. Bazylev, G. F. Matthews, S. Jachmich, I. Balboa, M. Clever, R. Dejarnac, I. Coffey, Y. Corre, S. Devaux, L. Frassinetti, E. Gauthier, J. Horacek, M. Knaup, M. Komm, K. Krieger, S. Marsen, A. Meigs, P. Mertens, R. A. Pitts, T. Puetterich, M. Rack, M. Stamp, G. Sergienko, P. Tamain, and V. Thompson, "ELM induced tungsten melting and its impact on tokamak operation," *Journal of Nuclear Materials*, vol. 463, pp. 78–84, 2015.

[22] G. Federici, A. Loarte, and G. Strohmayer, "Assessment of erosion of the ITER divertor targets during type I ELMs," *Plasma Physics and Controlled Fusion*, vol. 45, no. 9, pp. 1523–1547, 2003.

[23] N. Klimov, V. Podkovyrov, A. Zhitlukhin, D. Kovalenko, B. Bazylev, G. Janeschitz, I. Landman, S. Pestchanyi, G. Federici, A. Loarte, M. Merola, J. Linke, T. Hirai, and J. Compan, "Experimental study of PFCs erosion under ITER-like transient loads at plasma gun facility QSPA," *Journal of Nuclear Materials*, vol. 390-391, no. 1, pp. 721–726, 2009.

[24] D. Reiter, G. Wolf, and H. Kever, "Burn condition, helium particle confinement and exhaust efficiency," *Nuclear Fusion*, vol. 30, no. 10, p. 2141, 1990.

[25] A. C. C. Sips, for the Steady State Operation, and the Transport Physics topical groups of the International Tokamak Physics Activity, "Advanced scenarios for ITER operation," *Plasma Physics and Controlled Fusion*, vol. 47, no. 5A, pp. A19–A40, 2004.

[26] S. TheoDoridis and K. Koutroumbas, *Pattern Recognition*, 4th ed. Elsevier Inc., 2009.

[27] C. M. Bishop, *Pattern Recognition and Machine Learning.* Springer, 2006.

[28] A. Jain, R. Duin, and J. Mao, "Statistical pattern recognition: a review," pp. 4–37, 2000.

[29] R. Biswas, L. Blackburn, J. Cao, R. Essick, K. A. Hodge, E. Katsavounidis, K. Kim, Y. M. Kim, E. O. Le Bigot, C. H. Lee, J. J. Oh, S. H. Oh, E. J. Son, Y. Tao, R. Vaulin, and X. Wang, "Application of machine learning algorithms to the study of noise artifacts in gravitational-wave data," *Physical Review D - Particles, Fields, Gravitation and Cosmology*, vol. 88, no. 6, 2013.

[30] B. Krishnan, A. M. Sintes, M. A. Papa, B. F. Schutz, S. Frasca, and C. Palomba, "Hough transform search for continuous gravitational waves," *Physical Review D - Particles, Fields, Gravitation and Cosmology*, vol. 70, no. 8, pp. 1–23, 2004.

[31] A. Murari, J. Vega, G. Vagliasindi, J. A. Alonso, D. Alves, R. Coelho, S. DeFiore, J. Farthing, C. Hidalgo, and G. A. Rattá, "Recent developments in data mining and soft computing for JET with a view on ITER," *Fusion Engineering and Design*, vol. 84, no. 7-11, pp. 1372–1375, 2009.

[32] J. Vega, G. Rattá, P. Castro, A. Murari, and J.-E. contributors, "Development of learning systems with data tours techniques for fusion databases," in *World Scientific Proceedings Series on Computer Engineering and Information Science: Volume 1 Computational Intelligence in Decision and Control*, 2008, pp. 103–108.

[33] S.-I. Amari and H. Nagaoka, *Methods of Information Geometry.* Oxford University Press and the American Mathematical Society, 2000, vol. 191.

[34] L. Frassinetti, D. Dodt, M. Beurskens, A. Sirinelli, J. Boom, T. Eich, J. Flanagan, C. Giroud, M. Jachmich, M. Kempenaars, P. Lomas, G. Maddison, C. Maggi, R. Neu, I. Nunes, C. Perez von Thun, B. Sieglin, M. Stamp, and J.-E. contributors, "Effect of nitrogen seeding on the energy losses and on the time scales of the electron temperature and density collapse of type-I ELMs in JET with the ITER-like wall," *Nuclear Fusion*, vol. 55, no. 2, p. 023007, 2015.

[35] P. A. Schneider, L. B. Orte, A. Burckhart, M. G. Dunne, C. Fuchs, A. Gude, B. Kurzan, W. Suttrop, E. Wolfrum, and the ASDEX Upgrade Team, "Pedestal and edge localized mode characteristics with different first wall materials and nitrogen seeding in ASDEX Upgrade," *Plasma Physics and Controlled Fusion*, vol. 57, no. 1, p. 014029, 2015.

[36] S. Watanabe, *Pattern recognition: Human and mechanical*. New York: Wiley, 1985.

[37] M. Kirby, *Geometric Data Analysis*. John Wiley and Sons, 2001.

[38] S. Dormido-Canto, J. Vega, J. M. Ramírez, A. Murari, R. Moreno, J. M. López, A. Pereira, and J.-E. Contributors, "Development of an efficient real-time disruption predictor from scratch on JET and implications for ITER," *Nuclear Fusion*, vol. 53, no. 11, p. 113001, 2013.

[39] J. Vega, S. Dormido-Canto, J. M. López, A. Murari, J. M. Ramírez, R. Moreno, M. Ruiz, D. Alves, and R. Felton, "Results of the JET real-time disruption predictor in the ITER-like wall campaigns," in *Fusion Engineering and Design*, vol. 88, no. 6-8, 2013, pp. 1228–1231.

[40] B. Cannas, A. Fanni, A. Murari, A. Pau, and G. Sias, "Automatic disruption classification based on manifold learning for real-time applications on JET," *Nuclear Fusion*, vol. 53, no. 9, p. 093023, 2013.

[41] B. Cannas, P. C. de Vries, A. Fanni, A. Murari, A. Pau, and G. Sias, "Automatic disruption classification in JET with the ITER-like wall," *Plasma Physics and Controlled Fusion*, vol. 57, no. 12, p. 125003, 2015.

[42] E. J. Doyle, W. A.Houlberg, Y. Kamada, V. Mukhovatov, T. H.Osborne, A. Polevoi, G. Bateman, J. W.Connor, J. G.Cordey, T. Fujita, X. Garbet, T. S.Hahm, L. D.Horton, A. E.Hubbard, F. Imbeaux, F. Jenko, J. E.Kinsey, Y. Kishimoto, J. Li, T. C.Luce, Y. Martin, M. Ossipenko, V. Parail, A. Peeters, T. L.Rhodes, J. E.Rice, C. M.Roach, V. Rozhansky, F. Ryter, G. Saibene, R. Sartori, A. C.C.Sips, J. A.Snipes, M. Sugihara, E. J.Synakowski, H. Takenaga, T. Takizuka, K. Thomsen, M. R.Wade, H. R.Wilson, ITPA Transport Physics Topical Group, ITPA Confinement Database and Modelling Topical Group, and ITPA Pedestal and Edge Topical Group, "Chapter 2: Plasma confinement and transport," *Nuclear Fusion*, vol. 47, no. 6, p. S18, 2007.

[43] A. Murari, E. Peluso, M. Lungaroni, M. Gelfusa, and P. Gaudio, "Application of symbolic regression to the derivation of scaling laws for tokamak energy confinement time in terms of dimensionless quantities," *Nuclear Fusion*, vol. 56, no. 2, p. 026005, 2015.

[44] G. Verdoolaege and J. M. Noterdaeme, "Robust scaling in fusion science: Case study for the L-H power threshold," *Nuclear Fusion*, vol. 55, no. 11, p. 113019, 2015.

[45] M. M. Deza and E. Deza, *Encyclopedia of distances*. Springer-Verlag, 2009.

[46] A. Berman and L. Shapiro, "A flexible image database system for content-based retrieval," *Proceedings. Fourteenth International Conference on Pattern Recognition (Cat. No.98EX170)*, vol. 1, pp. 175–195, 1998.

[47] W. A. Burkhard and R. M. Keller, "Some approaches to best-match file searching," *Communications of the ACM*, vol. 16, no. 4, pp. 230–236, 1973.

[48] G. Verdoolaege and P. Scheunders, "Geodesics on the manifold of multivariate generalized Gaussian distributions with an application to multicomponent texture discrimination," *International Journal of Computer Vision*, vol. 95, no. 3, pp. 265–286, 2011.

[49] C. A. Castano-Moraga, C. Lenglet, R. Deriche, and J. Ruiz-Alzola, "A

Riemannian approach to anisotropic filtering of tensor fields," *Signal Processing*, vol. 87, no. 2, pp. 263–276, 2007.

[50] C. Lenglet, M. Rousson, and R. Deriche, "DTI segmentation by statistical surface evolution," *IEEE Transactions on Medical Imaging*, vol. 25, no. 6, pp. 685–700, 2006.

[51] O. Calin and C. Udriste, *Geometric modeling in probability and statistics*. Springer, 2014.

[52] H. Cramér, "A contribution to the theory of statistical estimation," *Skand. Aktuarietidskr.*, vol. 29, pp. 85–94, 1946.

[53] C. Radhakrishna Rao, "Information and accuracy attainable in the estimation of statistical parameters," *Bulletin of the Calcutta Mathematical Society*, vol. 37, no. 3, pp. 81–91, 1945.

[54] N. N.Cenkov, *Statistical Decision Rules and Optimal Inference (Translations of Mathematical Monographs)*. American Mathematical Society, 1982.

[55] G. Verdoolaege and P. Scheunders, "On the geometry of multivariate generalized Gaussian models," *Journal of Mathematical Imaging and Vision*, vol. 43, no. 3, pp. 180–193, 2012.

[56] J. Burbea and C. R. Rao, "Entropy differential metric, distance and divergence measures in probability spaces: A unified approach," *Journal of Multivariate Analysis*, vol. 12, no. 4, pp. 575–596, 1982.

[57] S. G. Mallat, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, 1989.

[58] S. G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression." *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 9, no. 9, pp. 1532–1546, 2000.

[59] M. N. Do and M. Vetterli, "Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance," *IEEE Transactions on Image Processing*, vol. 11, no. 2, pp. 146–158, 2002.

[60] S. G. Chang, B. Yu, and M. Vetterli, "Spatially adaptive wavelet thresholding with context modeling for image denoising." *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 9, no. 9, pp. 1522–31, 2000.

[61] J. Scharcanski, "Stochastic texture analysis for monitoring stochastic processes in industry," *Pattern Recognition Letters*, vol. 26, no. 11, pp. 1701–1709, 2005.

[62] G. Verdoolaege, G. Karagounis, A. Murari, J. Vega, and G. Van Oost, "Modeling fusion data in probabilistic metric spaces: Applications to the identification of confinement regimes and plasma disruptions," in *Fusion Science and Technology*, vol. 62, no. 2, 2012, pp. 356–365.

[63] P. Scheunders and S. De Backer, "Wavelet denoising of multicomponent images using Gaussian scale mixture models and a noise-free Image as priors," *IEEE Transactions on Image Processing*, vol. 16, no. 7, pp. 1865–1872, 2007.

[64] J. M. Oller, "Information Metric for Extreme Value and Logistic Probability Distributions," *Sankhya, The Indian Journal of Statistics, Series A*, vol. 49, no. 1, pp. 17–23, 1987.

[65] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 1990, vol. 22.

[66] R. E. Bellman, *Adaptive Control Processes: A Guided Tour*. Princeton University Press, 1961.

[67] D. W. Scott, *Multivariate Density Estimation: Theory, Practice, and Visualization (Wiley Series in Probability and Statistics)*, 1992, vol. 156.

[68] S. J. Raudys and A. K. Jain, "Small sample size effects in statistical pattern recognition: recommendations for practitioners," pp. 252–264, 1991.

[69] G. A. Ratta, J. Vega, A. Murari, and . JET-EFDA contributors, "Improved feature selection based on genetic algorithms for real time disruption prediction on JET," *Fusion Engineering and Design*, vol. 87, no. 9, pp. 1670–1678, 2012.

[70] A. Pereira, J. Vega, R. Moreno, S. Dormido-Canto, G. A. Ratta, F. Pavon, and . JET-EFDA contributors, "Feature selection for disruption prediction from scratch in JET by using genetic algorithms and probabilistic predictors," *Fusion Engineering and Design*, vol. 96-97, pp. 907–911, 2015.

[71] J. Parkkinen, K. Nybo, J. Peltonen, and S. Kaski, "Graph visualization with latent variable models," in *International Conference on Knowledge Discovery and Data Mining*, 2010, pp. 94–101.

[72] M. Camplani, B. Cannas, A. Fanni, G. Pautasso, and G. Sias, "Tracking of the plasma states in a nuclear fusion device using SOMs," *Neural Computing and Applications*, vol. 20, no. 6, pp. 851–863, 2011.

[73] C. M. Bishop, M. Svensdn, and C. K. I. Williams, "GTM: A principled alternative to the Self-Organizing Map," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 1112 LNCS, 1996, pp. 165–170.

[74] T. Kohonen, "The self-organizing map," *Neurocomputing*, vol. 21, no. 1-3, pp. 1–6, 1998.

[75] S. Kaski and J. Peltonen, "Dimensionality Reduction for Data Visualization," *IEEE Signal Processing Magazine*, vol. 28, no. 2, pp. 100–104, 2011.

[76] T. F. Cox and M. A. A. Cox, *Multidimensional Scaling*. Chapman and Hall, 2001.

[77] J. B. Kruskal and M. Wish, *Multidimensional Scaling*, 1978, vol. 31.

[78] R. N. Shepard, "Multidimensional scaling, tree-fitting, and clustering." pp. 390–398, 1980.

[79] V. D. Silva and J. B. Tenenbaum, "Sparse multidimensional scaling using landmark points," *Technology*, vol. 33, no. 6, pp. 1–41, 2004.

[80] A. Shabbir, G. Verdoolaege, and G. Van Oost, "Multivariate texture discrimination based on geodesics to class centroids on a generalized Gaussian manifold," in *Geometric Science of Information Volume 8085 of the series Lecture Notes in Computer Science*, 2013, pp. 853–860.

[81] G. Verdoolaege, S. De Backer, and P. Scheunders, "Multiscale colour texture retrieval using the geodesic distance between multivariate generalized gaussian models," in *Proceedings - International Conference on Image Processing, ICIP*, 2008, pp. 169–172.

[82] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1967.

[83] E. Fix and J. Hodges Jr., "Discriminatory analysis, Nonparametric discrimination: consistency properties," USAF School of Aviation Medicine Randolph Field, Texas, Tech. Rep., 1951.

[84] K. Weinberger, J. Blitzer, and L. Saul, "Distance metric learning for large margin nearest neighbor classification," *Advances in neural information processing systems*, vol. 18, p. 1473, 2006.

[85] X. Pennec, "Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements," *Journal of Mathematical Imaging and Vision*, vol. 25, no. 1, pp. 127–154, 2006.

[86] A. Laine and J. Fan, "Texture Classification by Wavelet Packet Signatures," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 11, pp. 1186–1191, 1993.

[87] B. S. Manjunath, "Texture features for browsing and retrieval of image data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 8, pp. 837–842, 1996.

[88] J. G. Daugman, "Two-dimensional spectral analysis of cortical receptive field profiles," *Vision Research*, vol. 20, no. 10, pp. 847–856, 1980.

[89] G. Verdoolaege, G. Karagounis, M. Tendler, and G. V. Oost, "Pattern recognition in probability spaces for visualization and identification of plasma confinement regimes and confinement time scaling," *Plasma Physics and Controlled Fusion*, vol. 54, no. 12, p. 124006, 2012.

[90] X. Pennec, P. Fillard, and N. Ayache, "A riemannian framework for tensor computing," *International Journal of Computer Vision*, vol. 66, no. 1, pp. 41–66, 2006.

[91] "MIT vision and modeling group, Vision texture," http://vismod.media.mit.edu/vismod/imagery/VisionTexture/., 2010.

[92] "The Mathworks," www.mathworks.com., 2012.

[93] R. Kwitt and A. Uhl, "Lightweight probabilistic texture retrieval," *IEEE Transactions on Image Processing*, vol. 19, no. 1, pp. 241–253, 2010.

[94] G. Verdoolaege, Y. Rosseel, M. Lambrechts, and P. Scheunders, "Wavelet-based colour texture retrieval using the Kullback-Leibler divergence between bivariate generalized Gaussian models," in *Proceedings - International Conference on Image Processing, ICIP*, 2009, pp. 265–268.

[95] I. T. Jolliffe, *Principal Component Analysis*. Springer-Verlag, 1986.

[96] P. T. Fletcher, C. Lu, S. M. Pizer, and S. Joshi, "Principal geodesic analysis for the study of nonlinear statistics of shape," in *IEEE Transactions on Medical Imaging*, vol. 23, no. 8, 2004, pp. 995–1005.

[97] X. Pennec, "Probabilities and Statistics on Riemannian Manifolds: Basic Tools for Geometric Measurements," in *IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing*, 1999.

[98] A. W. Leonard, "Edge-localized-modes in tokamaks," *Physics of Plasmas*, vol. 21, no. 9, 2014.

[99] W. Suttrop, "The physics of large and small edge localized modes," *Plasma Physics and Controlled Fusion*, vol. 42, no. 5A, pp. A1–A14, 2000.

[100] A. W. Leonard, A. Herrmann, K. Itami, J. Lingertat, A. Loarte, T. H. Osborne, and W. Suttrop, "Impact of ELMs on the ITER divertor," *Journal of Nuclear Materials*, vol. 266, pp. 109–117, 1999.

[101] E. J. Doyle, R. J. Groebner, K. H. Burrell, P. Gohil, T. Lehecka, N. C. Luhmann, H. Matsumoto, T. H. Osborne, W. A. Peebles, and R. Philipona, "Modifications in Turbulence and Edge Electric-Fields at the L-H Transition in the DIII-D Tokamak," *Physics of Fluids B-Plasma Physics*, vol. 3, no. 8, pp. 2300–2307, 1991.

[102] A. J. Webster, R. O. Dendy, and J.-E. contributors, "Supplementary Material for Statistical Characterisation and Classification of Ege Localised Plasma Instabilities," *Physical Review Letters*, vol. 110, no. 15, p. 155004, 2013.

[103] A. Loarte, G. Saibene, R. Sartori, D. Campbell, M. Becoulet, L. Horton, T. Eich, A. Herrmann, G. Matthews, N. Asakura, A. Chankin, A. Leonard, G. Porter, G. Federici, G. Janeschitz, M. Shimada, and M. Sugihara, "Characteristics of type I ELM energy and particle losses in existing devices and their extrapolation to ITER," *Plasma Physics and Controlled Fusion*, vol. 45, no. 9, pp. 1549–1569, 2003.

[104] D. Zarzoso, M. Beurskens, L. Frassinetti, E. Joffrin, F. Rimini, E. Solano, and JET EFDA contributors, "ELM size analysis in JET hybrid plasmas," *Nuclear Fusion*, vol. 51, no. 11, p. 112001, 2011.

[105] . ITER Physics Expert Group on Confinement and Transport, ITER Physics Expert Group on Confinement Modelling and Database and ITER Physics Basis Editors, "Chapter 2 : Plasma confinement and transport," *Nuclear Fusion*, vol. 39, no. 12, 1999.

[106] L. Horton, J. Christiansen, J. Lingertat, C. Maggi, V. Mertens, O. Pogutse, G. Saibene, R. Sartori, J. Stober, W. Suttrop, T. J. Team, and the ASDEX Upgrade Team, "Performance near operational boundaries," *Plasma Physics and Controlled Fusion*, vol. 329, no. 41, pp. B329–B341, 1999.

[107] A. Loarte, M. Becoulet, G. Saibene, R. Sartori, D. J. Campbell, T. Eich, A. Herrmann, M. Laux, W. Suttrop, B. Alper, P. J. Lomas, G. Matthews, S. Jachmich, J. Ongena, P. Innocente, and EFDA-JET Workprogramme Collaborators, "Characteristics and scaling of energy and particle losses during Type I ELMs in JET H-modes," *Plasma Physics and Controlled Fusion*, vol. 44, no. 9, pp. 1815–1844, sep 2002.

[108] A. J. Webster, S. J. Webster, and J.-E. contributors, "Processes and properties of edge-localised instabilities in 2T 2MA plasmas in the Joint European Torus," *Physics of Plasmas*, vol. 21, no. 11, p. 112502, 2014.

[109] J. Stober, M. Maraschek, G. Conway, O. Gruber, A. Herrmann, A. Sips, W. Treutterer, H. Zohm, and ASDEX Upgrade Team, "Type II ELMy H modes on ASDEX Upgrade with good confinement at high density," *Nuclear Fusion*, vol. 41, no. 9, pp. 1123–1134, 2001.

[110] G. Saibene, P. Lomas, R. Sartori, A. Loarte, J. Stober, Y. Andrew, S. Arshad, G. Conway, E. D. L. Luna, K. Günther, L. Ingesson, M. Kempenaars, A. Korotkov, H. Koslowski, J. Lönnroth, D. McDonald, A. Meigs, P. Monier-Garbet, V. Parail, C. Perez, F. Rimini, S. Sharapov, and P. Thomas, "Characterization of small ELM experiments in highly shaped single null and quasi-double-null plasmas in JET," *Nuclear Fusion*, vol. 45, no. 5, pp. 297–317, 2005.

[111] J. Connor, A. Kirk, and H. Wilson, "Edge Localised Modes (ELMs): Experiment and Theory," *AIP Conference Proceedings*, vol. 1013, pp. 174–190, 2008.

[112] A. Kirk, W. Suttrop, I. Chapman, Y. Liu, R. Scannell, A. Thornton, L. B. Orte, P. Cahyna, T. Eich, R. Fischer, C. Fuchs, C. Ham, J. Harrison, M. Jakubowski, B. Kurzan, S. Pamela, M. Peterka, D. Ryan, S. Saarelma, B. Sieglin, M. Valovic, M. Willensdorfer, and MAST and ASDEX Upgrade Team, "Effect of resonant magnetic perturbations on low collisionality discharges in MAST and a comparison with ASDEX Upgrade," *Nuclear Fusion*, vol. 55, no. 4, p. 043011, 2015.

[113] R. Maingi, K. Tritz, E. D. Fredrickson, J. E. Menard, S. A. Sabbagh, D. Stutman, M. G. Bell, R. E. Bell, C. E. Bush, D. A. Gates, D. W. Johnson, R. Kaita, S. M. Kaye, H. W. Kugel, B. P. LeBlanc, D. Mueller, R. Raman, A. L. Roquemore, and V. A. Soukhanovskii, "Observation of a high performance operating regime with small edge-localized modes in the National Spherical Torus Experiment," *Nuclear Fusion*, vol. 45, no. 4, pp. 264–270, 2005.

[114] R. Maingi, M. G. Bell, E. D. Fredrickson, K. C. Lee, R. J. Maqueda, P. Snyder, K. Tritz, S. J. Zweben, R. E. Bell, T. M. Biewer, C. E. Bush, J. Boedo, N. H. Brooks, L. Delgado-Aparicio, C. W. Domier, D. A. Gates, D. W. Johnson, R. Kaita, S. M. Kaye, H. W. Kugel, B. P. LeBlanc, N. C. Luhmann, J. E.

Menard, D. Mueller, H. Park, R. Raman, A. L. Roquemore, S. A. Sabbagh, V. A. Soukhanovskii, T. Stevenson, and D. Stutman, "Characterization of small, Type V edge-localized modes in the National Spherical Torus Experiment," *Physics of Plasmas*, vol. 13, no. 9, 2006.

[115] Y. Kamada, T. Oikawa, L. Lao, T. Takizuka, T. Hatae, A. Isayama, J. Manickam, M. Okabayashi, T. Fukuda, and K. Tsuchiya, "Disappearance of giant ELMs and appearance of minute grassy ELMs in JT-60U high-triangularity discharges," *Plasma Physics and Controlled Fusion*, vol. 42, no. 5A, pp. A247–A253, 2000.

[116] N. Oyama, Y. Sakamoto, A. Isayama, M. Takechi, P. Gohil, L. Lao, P. Snyder, T. Fujita, S. Ide, Y. Kamada, Y. Miura, T. Oikawa, T. Suzuki, H. Takenaga, K. Toi, and the JT-60 Team, "Energy loss for grassy ELMs and effects of plasma rotation on the ELM characteristics in JT-60U," *Nuclear Fusion*, vol. 45, no. November, pp. 871–881, 2005.

[117] Y. Kamada, H. Takenaga, A. Isayama, T. Hatae, H. Urano, H. Kubo, T. Takizuka, and Y. Miura, "Pedestal characteristics and extended high-beta(p) ELMy H-mode regime in JT-60U," *Plasma Physics and Controlled Fusion*, vol. 44, pp. A279–A286, 2002.

[118] J. Stober, P. J. Lomas, G. Saibene, Y. Andrew, P. Belo, G. D. Conway, A. Herrmann, L. D. Horton, M. Kempenaars, H.-R. R. Koslowski, A. Loarte, G. P. Maddison, M. Maraschek, D. C. McDonald, a. G. Meigs, P. Monier-Garbet, D. A. Mossessian, M. F. F. Nave, N. Oyama, V. Parail, C. P. Perez, F. Rimini, R. Sartori, A. Sips, P. Thomas, C. t. t. E.-J. Workprogramme, and the ASDEX Upgrade Team, "Small ELM regimes with good confinement on JET and comparison to those on ASDEX upgrade, Alcator C-mod and JT-60U," *Nuclear Fusion*, vol. 45, no. 11, pp. 1213–1223, 2005.

[119] P. Lang, A. Loarte, G. Saibene, L. Baylor, M. Becoulet, M. Cavinato, S. Clement-Lorenzo, E. Daly, T. Evans, M. Fenstermacher, Y. Gribov, L. Horton, C. Lowry, Y. Martin, O. Neubauer, N. Oyama, M. Schaffer, D. Stork, W. Suttrop, P. Thomas, M. Tran, H. Wilson, A. Kavin, and

O. Schmitz, "ELM control strategies and tools: status and potential for ITER," *Nuclear Fusion*, vol. 53, no. 4, p. 043004, 2013.

[120] D. A. Mossessian, P. B. Snyder, M. Greenwald, J. W. Hughes, Y. Lin, A. Mazurenko, S. Medvedev, H. R. Wilson, and S. Wolfe, "H-mode pedestal characteristics and MHD stability of the edge plasma in Alcator C-Mod," *Analysis*, vol. 44, pp. 423–437, 2002.

[121] Y. Takase, R. L. Boivin, F. Bombarda, P. T. Bonoli, C. Christensen, C. Fiore, D. T. Garnier, J. A. Goetz, S. N. Golovato, R. S. Granetz, M. J. Greenwald, S. F. Horne, A. E. Hubbard, I. H. Hutchinson, J. H. Irby, B. LaBombard, B. Lipschultz, E. S. Marmar, M. J. May, A. Mazurenko, G. M. McCracken, P. OShea, M. Porkolab, J. Reardon, J. E. Rice, C. J. Rost, J. Schachter, J. A. Snipes, P. C. Stek, J. L. Terry, R. L. Watterson, B. Welch, and S. M. Wolfe, "Radiofrequency-heated enhanced confinement modes in the Alcator C-Mod tokamak," *Phys. Plasmas*, vol. 4, pp. 1647–1653, 1997.

[122] K. Kamiya, H. Kimura, H. Ogawa, H. Kawashima, K. Tsuzuki, M. Sato, Y. Miura, and JFT-2M Group, "Observation of High Recycling Steady H-mode edge and compatibility with improved core confinement mode on JFT-2M," *Nuclear Fusion*, vol. 43, no. 10, pp. 1214–1219, 2003.

[123] K. H. Burrell, A. M. Garofalo, W. M. Solomon, M. E. Fenstermacher, T. H. Osborne, J. K. Park, M. J. Schaffer, and P. B. Snyder, "Reactor-relevant quiescent H-mode operation using torque from non-axisymmetric, non-resonant magnetic fields," *Physics of Plasmas*, vol. 19, no. 5, 2012.

[124] D. Whyte, A. Hubbard, J. Hughes, B. Lipschultz, J. Rice, E. Marmar, M. Greenwald, I. Cziegler, A. Dominguez, T. Golfinopoulos, N. Howard, L. Lin, R. McDermott, M. Porkolab, M. Reinke, J. Terry, N. Tsujii, S. Wolfe, S. Wukitch, Y. Lin, and the Alcator C-Mod Team, "I-mode: an H-mode energy confinement regime with L-mode particle transport in Alcator C-Mod," *Nuclear Fusion*, vol. 50, no. 10, p. 105005, 2010.

[125] A. E. Hubbard, D. G. Whyte, R. M. Churchill, I. Cziegler, A. Dominguez, T. Golfinopoulos, J. W. Hughes, J. E. Rice, I. Bespamyatnov, M. J. Greenwald,

N. Howard, B. Lipschultz, E. S. Marmar, M. L. Reinke, W. L. Rowan, and J. L. Terry, "Edge energy transport barrier and turbulence in the I-mode regime on alcator C-mod," *Physics of Plasmas*, vol. 18, no. 5, 2011.

[126] M. Keilhacker, G. V. Gierke, E. R. Muller, H. Murmann, F. Soldner, E. Speth, K. Steinmetz, F. Wagner, R. Bartiromo, G. Becker, H.-S. Bosch, A. Eberhagen, D. Eckhartt, G. Fussmann, O. Gehre, J. Gernhardt, A. Giuliana, E. Glock, O. Gruber, G. Haas, G. Janeschitz, F. Karger, O. Kluber, M. Kornherr, K. Lackner, M. Lenoci, F. Leuterer, G. Lisitano, H. M. Mayer, K. McCormick, D. Meisel, V. Mertens, M. Munich, H. Niedermeyer, W. Poschenrieder, H. Rapp, F. Ryter, F. Schneider, G. Siller, P. Smeulders, K. H. Steuer, O. Vollmer, F. Wesner, and D. Zasche, "Confinement in ASDEX with neutral beam and RF heating," *Plasma Physics and Controlled Fusion*, vol. 28, pp. 29–41, 1986.

[127] Y. Liang, "Edge Localised Mode (ELM)," in *Active Control of Magneto-hydrodynamic Instabilities in Hot Plasmas*, I. Valentin, Ed. Springer, 2015, pp. 143–182.

[128] J. Boom, *Characterization of edge localised modes in tokamak plasmas.* PhD Dissertation,Technical University Eindhoven, 2012.

[129] J. Connor, "A review of models for ELMs," *Plasma Physics and Controlled Fusion*, vol. 40, pp. 191–213, 1998.

[130] P. B. Snyder, H. R. Wilson, J. R. Ferron, L. L. Lao, A. W. Leonard, T. H. Osborne, A. D. Turnbull, D. Mossessian, M. Murakami, and X. Q. Xu, "Edge localized modes and the pedestal: A model based on coupled peeling–ballooning modes," *Physics of Plasmas*, vol. 9, no. 5, p. 2037, 2002.

[131] H. R. Wilson and R. L. Miller, "Access to second stability region for coupled peeling-ballooning modes in tokamaks," *Physics of Plasmas*, vol. 6, no. 3, p. 873, 1999.

[132] L. R. Baylor, N. Commaux, T. C. Jernigan, N. H. Brooks, S. K. Combs, T. E. Evans, M. E. Fenstermacher, R. C. Isler, C. J. Lasnier, S. J. Meitner, R. A. Moyer, T. H. Osborne, P. B. Parks, P. B. Snyder, E. J. Strait, E. A.

Unterberg, and A. Loarte, "Reduction of edge-localized mode intensity using high-repetition-rate pellet injection in tokamak H-mode plasmas," *Physical Review Letters*, vol. 110, no. 24, 2013.

[133] W. Xiao, P. Diamond, X. Zou, J. Dong, X. Ding, L. Yao, B. Feng, C. Chen, W. Zhong, M. Xu, B. Yuan, T. Rhee, J. Kwon, Z. Shi, J. Rao, G. Lei, J. Cao, J. Zhou, M. Huang, D. Yu, Y. Huang, K. Zhao, Z. Cui, X. Song, Y. Gao, Y. Zhang, J. Cheng, X. Han, Y. Zhou, Y. Dong, X. Ji, Q. Yang, Y. Liu, L. Yan, X. Duan, and Y. Liu, "ELM mitigation by supersonic molecular beam injection into the H-mode pedestal in the HL-2A tokamak," *Nuclear Fusion*, vol. 52, no. 11, p. 114027, 2012.

[134] J. S. Hu, Z. Sun, C. Z. Li, X. W. Zhen, J. G. Li, H. Y. Guo, J. H. Li, L. Wang, K. F. Gan, Y. Chen, J. Ren, G. Z. Zuo, X. J. Yao, L. Q. Hu, X. Z. Gong, B. N. Wan, X. L. Zou, D. K. Mansfield, Y. F. Liang, and I. Vinyar, "ELM mitigation by means of supersonic molecular beam and pellet injection on the EAST superconducting tokamak," *Journal of Nuclear Materials*, vol. 463, pp. 718–722, 2015.

[135] E. de la Luna, I. Chapman, F. Rimini, P. Lomas, G. Saibene, F. Koechl, R. Sartori, S. Saarelma, R. Albanese, J. Flanagan, F. Maviglia, V. Parail, A. Sips, E. Solano, and J. contributors, "Understanding the physics of ELM pacing via vertical kicks in JET in view of ITER," *Nuclear Fusion*, vol. 56, no. 2, p. 026001, 2016.

[136] J. M. Canik, R. Maingi, T. E. Evans, R. E. Bell, S. P. Gerhardt, B. P. Leblanc, J. Manickam, J. E. Menard, T. H. Osborne, J. K. Park, S. F. Paul, P. B. Snyder, S. A. Sabbagh, H. W. Kugel, and E. A. Unterberg, "On demand triggering of Edge localized instabilities using external nonaxisymmetric magnetic perturbations in toroidal plasmas," *Physical Review Letters*, vol. 104, no. 4, 2010.

[137] Y. Liang, H. R. Koslowski, P. R. Thomas, E. Nardon, S. Jachmich, B. Alper, P. Andrew, Y. Andrew, G. Arnoux, Y. Baranov, M. Becoulet, M. Beurskens, T. Biewer, M. Bigi, K. Crombe, E. De La Luna, P. de Vries, T. Eich, H. G. Esser, W. Fundamenski, S. Gerasimov, C. Giroud, M. P. Gryaznevich,

D. Harting, N. Hawkes, S. Hotchin, D. Howell, A. Huber, M. Jakubowski, V. Kiptily, A. Kreter, L. Moreira, V. Parail, S. D. Pinches, E. Rachlew, O. Schmitz, O. Zimmermann, and J.-F. Contributors, "Active control of type-I edge localized modes on JET," *Plasma Physics and Controlled Fusion*, vol. 49, no. 12B, pp. B581–B589, 2007.

[138] R. A. Moyer, T. E. Evans, T. H. Osborne, P. R. Thomas, M. Becoulet, J. Harris, K. H. Finken, J. A. Boedo, E. J. Doyle, M. E. Fenstermacher, P. Gohil, R. J. Groebner, M. Groth, G. L. Jackson, R. J. La Haye, C. J. Lasnier, A. W. Leonard, G. R. McKee, H. Reimerdes, T. L. Rhodes, D. L. Rudakov, M. J. Schaffer, P. B. Snyder, M. R. Wade, G. Wang, J. G. Watkins, W. P. West, and L. Zeng, "Edge localized mode control with an edge resonant magnetic perturbation," in *Physics of Plasmas*, vol. 12, no. 5, 2005, pp. 1–11.

[139] A. Herrmann, "Overview on stationary and transient divertor heat loads," *Plasma Physics and Controlled Fusion*, vol. 44, no. 6, pp. 883–903, 2002.

[140] T. Eich, H. Thomsen, W. Fundamenski, G. Arnoux, S. Brezinsek, S. Devaux, A. Herrmann, S. Jachmich, J. Rapp, and J.-E. contributors, "Type-I ELM power deposition profile width and temporal shape in JET," in *Journal of Nuclear Materials*, vol. 415, no. 1 SUPPL, 2011.

[141] H. Thomsen, T. Eich, S. Devaux, G. Arnoux, S. Brezinsek, E. DelaLuna, W. Fundamenski, A. Herrmann, A. Huber, S. Jachmich, P. Lomas, I. Nunes, G. Saibene, A. Scarabosio, J. Schweinzer, and J.-E. contributors, "Power load characterization for type-I ELMy H-modes in JET," *Nuclear Fusion*, vol. 51, no. 12, p. 123001, 2011.

[142] A. Loarte, G. Huijsmans, S. Futatani, L. Baylor, T. Evans, D. M. Orlov, O. Schmitz, M. Becoulet, P. Cahyna, Y. Gribov, A. Kavin, A. S. Naik, D. Campbell, T. Casper, E. Daly, H. Frerichs, A. Kischner, R. Laengner, S. Lisgo, R. Pitts, G. Saibene, and A. Wingen, "Progress on the application of ELM control schemes to ITER scenarios from the non-active phase to DT operation," *Nuclear Fusion*, vol. 54, no. 3, p. 033007, 2014.

[143] A. Loarte, D. Campbell, Y. Gribov, R. A. Pitts, N. Klimov, V. Podkovyrov, A. Zhitlukhin, I. Landman, S. Pestchanyi, B. Bazylev, J. Linke, T. Loewenhoff,

G. Pintsuk, O. Schmitz, Y. Liang, T. E. Evans, M. Schaffer, M. Fenstermacher, M. Becoulet, G. Huysmans, E. Nardon, L. Baylor, J. Canik, R. Maingi, J. Ahn, B. Riccardi, G. Saibene, R. Sartori, M. Cavinato, T. Eich, M. Jakubowski, P. Lang, H. Thomsen, W. Suttrop, E. de la Luna, H. Wilson, and A. Kirk, "ITER ELM control requirements, ELM control schemes & required R&D," in *23rd International Conference on Fusion Energy*, Daejeon, South Korea, 2010, pp. ITR/1–4.

[144] A. Bortolon, R. Maingi, D. K. Mansfield, A. Nagy, A. L. Roquemore, L. R. Baylor, N. Commaux, G. L. Jackson, E. P. Gilson, R. Lunsford, P. B. Parks, C. Chrystal, B. A. Grierson, R. Groebner, S. Haskey, M. J. Makowski, C. J. Lasnier, R. Nazikian, T. Osborne, D. Shiraki, and M. A. V. Zeeland, "High frequency pacing of edge localized modes by injection of lithium granules in DIII-D H-mode discharges," *Nuclear Fusion*, vol. 56, p. 056008, 2016.

[145] R. Neu, G. Arnoux, M. Beurskens, V. Bobkov, S. Brezinsek, J. Bucalossi, G. Calabro, C. Challis, J. W. Coenen, E. De La Luna, P. C. De Vries, R. Dux, L. Frassinetti, C. Giroud, M. Groth, J. Hobirk, E. Joffrin, P. Lang, M. Lehnen, E. Lerche, T. Loarer, P. Lomas, G. Maddison, C. Maggi, G. Matthews, S. Marsen, M. L. Mayoral, A. Meigs, P. Mertens, I. Nunes, V. Philipps, T. Pütterich, F. Rimini, M. Sertoli, B. Sieglin, A. C. C. Sips, D. Van Eester, G. Van Rooij, and J.-E. contributors, "First operation with the JET International Thermonuclear Experimental Reactor-like wall," *Physics of Plasmas*, vol. 20, no. 5, 2013.

[146] R. Dux, A. Janzer, T. Pütterich, and ASDEX Upgrade Team, "Main chamber sources and edge transport of tungsten in H-mode plasmas at ASDEX Upgrade," *Nuclear Fusion*, vol. 51, no. 5, p. 053002, 2011.

[147] "International Tokamak Physics Activity H-mode Database Working Group," http://efdasql.ipp.mpg.de/hmodepublic (2007).

[148] A. Murari, J. Vega, D. Mazon, G. A. Rattà, J. Svensson, S. Palazzo, G. Vagliasindi, P. Arena, C. Boulbe, B. Faugeras, L. Fortuna, and D. Moreau, "Innovative signal processing and data analysis methods on JET for control in

the perspective of next-step devices," *Nuclear Fusion*, vol. 50, no. 5, p. 55005, 2010.

[149] A. Shabbir, G. Verdoolaege, O. J. W. F. Kardaun, A. J. Webster, R. O. Dendy, and J.-M. Noterdaeme, "Discrimination and visualization of ELM types based on a probabilistic description of inter-ELM waiting times," in *41st European Physical Society Conference on Plasma Physics*, 2014, p. P1.001.

[150] V. Vovk, A. Gammerman, and C. Saunders, "Machine-Learning Applications of Algorithmic Randomness," in *16th International Conference on Machine Learning*, 1999, pp. 444–453.

[151] V. Vovk, A. Gammerman, and G. Shafer, *Algorithmic learning in a random world*, 2005.

[152] M. Bécoulet, G. Huysmans, Y. Sarazin, X. Garbet, P. Ghendrih, F. Rimini, E. Joffrin, X. Litaudon, P. Monier-Garbet, J.-M. Ané, P. Thomas, A. Grosman, V. Parail, H. R. Wilson, P. Lomas, P. DeVries, K.-D. Zastrow, G. F. Matthews, J. Lonnroth, S. Gerasimov, S. Sharapov, M. P. Gryaznevich, G. F. Counsell, A. Kirk, M. Valovic, R. J. Buttery, A. Loarte, G. Saibene, R. Sartori, A. Leonard, P. B. Snyder, L. L. Lao, P. Gohil, T. E. Evans, R. A. Moyer, Y. Kamada, A. Chankin, N. Oyama, T. Hatae, N. Asakura, O. Tudisco, E. Giovannozzi, F. Crisanti, C. P. Perez, H. R. Koslowski, T. Eich, A. Sips, L. Horton, A. Hermann, P. Lang, J. Stober, W. Suttrop, P. Beyer, S. Saarelma, and Contributors to JET-EFDA Work programme, "Edge localized mode physics and operational aspects in tokamaks," *Plasma Physics and Controlled Fusion*, vol. 45, no. 12A, pp. A93–A113, 2003.

[153] H. R. Wilson, P. B. Snyder, G. T. A. Huysmans, and R. L. Miller, "Numerical studies of edge localized instabilities in tokamaks," *Physics of Plasmas*, vol. 9, no. 4, p. 1277, 2002.

[154] P. Snyder, H. Wilson, J. R. Ferron, L. L. Lao, A. W. Leonard, D. Mossessian, M. Murakami, T. H. Osborne, A. D. Turnbull, and X. Q. Xu, "ELMs and constraints on the H-mode pedestal: peeling–ballooning stability calculation and comparison with experiment," *Nuclear Fusion*, vol. 44, no. 2, 2004.

[155] I. physics basis Editors, "ITER Physics Basis," *Nuclear Fusion*, vol. 39, no. 12, p. 2137, 1999.

[156] D. C. McDonald, J. G. Cordey, K. Thomsen, O. J. W. F. Kardaun, J. A. Snipes, M. Greenwald, L. Sugiyama, F. Ryter, A. Kus, J. Stober, J. C. DeBoo, C. C. Petty, G. Bracco, M. Romanelli, Z. Cui, Y. Liu, Y. Miura, K. Shinohara, K. Tsuzuki, Y. Kamada, T. Takizuka, H. Urano, M. Valovic, R. Akers, C. Brickley, A. Sykes, M. J. Walsh, S. M. Kaye, C. Bush, D. Hogewei, Y. R. Martin, A. Cote, G. Pacher, J. Ongena, F. Imbeaux, G. T. Hoang, S. Lebedev, A. Chudnovskiy, and V. Leonov, "Recent progress on the development and analysis of the ITPA global H-mode confinement database," *Nuclear Fusion*, vol. 47, no. 3, pp. 147–174, 2007.

[157] A. Shabbir, G. Verdoolaege, O. J. W. F. Kardaun, and J.-M. Noterdaeme, "Visualization of the operational space of edge-localized modes through low-dimensional embedding of probability distributions," *Review of Scientific Instruments*, vol. 85, no. 11, p. 11E819, 2014.

[158] O. J. W. F. Kardaun, *Classical Methods of Statistics*. Springer-Verlag, 2005.

[159] A. J. Webster and R. O. Dendy, "Statistical characterization and classification of edge-localized plasma instabilities," *Physical Review Letters*, vol. 110, no. 15, pp. 1–5, 2013.

[160] K. Mardia, J. Kent, and J. Bibby, *Multivariate analysis (Probability and Mathematical Statistics)*, 1st ed. Academic Press, 1980.

[161] E. Jaynes, *Probability Theory: The Logic of Science*. Cambridge University Press, 2003.

[162] A. Murari, F. Pisano, J. Vega, B. Cannas, A. Fanni, S. Gonzalez, M. Gelfusa, and M. Grosso, "Extensive statistical analysis of ELMs on JET with a carbon wall," *Plasma Physics and Controlled Fusion*, vol. 56, no. 11, p. 114007, 2014.

[163] W. S. Cleveland, "Robust Locally Weighted Regression and Smoothing Scatterplots," *Journal of the American Statistical Association*, vol. 74, no. 368, pp. 829–836, 1979.

[164] H. Zohm, "Edge localized modes (ELMs)," *Plasma Physics and Controlled Fusion*, vol. 38, no. 2, pp. 105–128, 1996.

[165] T. Evans, "ELM mitigation techniques," *Journal of Nuclear Materials*, vol. 438, pp. S11–S18, 2013.

[166] M. W. Jakubowski, T. E. Evans, M. E. Fenstermacher, M. Groth, C. J. Lasnier, A. W. Leonard, O. Schmitz, J. G. Watkins, T. Eich, W. Fundamenski, R. A. Moyer, R. C. Wolf, L. B. Baylor, J. A. Boedo, K. H. Burrell, H. Frerichs, J. S. DeGrassie, P. Gohil, I. Joseph, S. Mordijck, M. Lehnen, C. C. Petty, R. I. Pinsker, D. Reiter, T. L. Rhodes, U. Samm, M. J. Schaffer, P. B. Snyder, H. Stoschus, T. Osborne, B. Unterberg, E. Unterberg, and W. P. West, "Overview of the results on divertor heat loads in RMP controlled H-mode plasmas on DIII-D," *Nuclear Fusion*, vol. 49, no. 9, p. 14, 2009.

[167] A. Kallenbach, R. Neu, R. Dux, H.-U. Fahrbach, J. Fuchs, L. Giannone, O. Gruber, A. Hermann, P. Lang, B. Lipschultz, C. Maggi, J. Neuhauser, V. Philipps, T. Puetterich, V. Rohde, J. Roth, G. Sergienko, A. Sips, and ASDEX Upgrade Team, "Tokamak operation with high-Z plasma facing components," *Plasma Physics and Controlled Fusion*, vol. 47, no. 12B, 2005.

[168] R. Dux, A. Janzer, T. Puetterich, and ASDEX Upgrade Team, "Main chamber sources and edge transport of tungsten in H-mode plasmas at ASDEX Upgrade," *Nuclear Fusion*, vol. 51, no. 5, p. 119501, 2011.

[169] P. Lang, J. Neuhauser, L. Horton, T. Eich, L. Fattorini, J. Fuchs, O. Gehre, A. Herrmann, P. Ignácz, M. Jakobi, S. Kálvin, M. Kaufmann, G. Kocsis, B. Kurzan, C. Maggi, M. Manso, M. Maraschek, V. Mertens, A. Mück, H. Murmann, R. Neu, I. Nunes, D. Reich, M. Reich, S. Saarelma, W. Sandmann, J. Stober, U. Vogl, and the ASDEX Upgrade Team, "ELM frequency control by continuous small pellet injection in ASDEX Upgrade," *Nuclear Fusion*, vol. 43, no. 10, pp. 1110–1120, 2003.

[170] Y. Liang, "Overview of Edge Localized Modes Control in Tokamak Plasmas," *Fusion Science and Technology*, vol. 59, no. 3, pp. 586–601, 2011.

[171] M. Beurskens, L. Frassinetti, C. Challis, C. Giroud, S. Saarelma, B. Alper, C. Angioni, P. Bilkova, C. Bourdelle, S. Brezinsek, P. Buratti, G. Calabro, T. Eich, J. Flanagan, E. Giovannozzi, M. Groth, J. Hobirk, E. Joffrin, M. Leyland, P. Lomas, E. de la Luna, M. Kempenaars, G. Maddison, C. Maggi, P. Mantica, M. Maslov, G. Matthews, M.-L. Mayoral, R. Neu, I. Nunes, T. Osborne, F. Rimini, R. Scannell, E. Solano, P. Snyder, I. Voitsekhovitch, and P. de Vries, "Global and pedestal confinement in JET with a Be/W metallic wall," *Nuclear Fusion*, vol. 54, no. 4, p. 043001, 2014.

[172] C. Giroud, G. Maddison, S. Jachmich, F. Rimini, M. Beurskens, I. Balboa, S. Brezinsek, R. Coelho, J. Coenen, L. Frassinetti, E. Joffrin, M. Oberkofler, M. Lehnen, Y. Liu, S. Marsen, K. McCormick, A. Meigs, R. Neu, B. Sieglin, G. van Rooij, G. Arnoux, P. Belo, M. Brix, M. Clever, I. Coffey, S. Devaux, D. Douai, T. Eich, J. Flanagan, S. Grünhagen, A. Huber, M. Kempenaars, U. Kruezi, K. Lawson, P. Lomas, C. Lowry, I. Nunes, A. Sirinnelli, A. Sips, M. Stamp, and S. Wiesen, "Impact of nitrogen seeding on confinement and power load control of a high-triangularity JET ELMy H-mode plasma with a metal wall," *Nuclear Fusion*, vol. 53, no. 11, p. 113025, 2013.

[173] Joseph Ngatchou-Wandji, "On Testing for the Nullity of Some Skewness Coefficients," *International Statistical Review*, vol. 74, no. 1, pp. 47–65, 2006.

[174] W. D. Johnson and J. D. Gibbons, *Nonparametric Statistical Inference*, 1973, vol. 15, no. 2.

[175] E.L. Lehmann and J. P. Romano, *Testing Statistical Hypothesis*, 3rd ed. Springer, 2005.

[176] A. Shabbir, G. Verdoolaege, J. Vega, and A. Murari, "ELM Regime Classification by Conformal Prediction on an Information Manifold," *IEEE Transactions on Plasma Science*, pp. 1–10, 2015.

[177] N. Oyama, Y. Sakamoto, A. Isayama, M. Takechi, P. Gohil, L. Lao, P. Snyder, T. Fujita, S. Ide, Y. Kamada, Y. Miura, T. Oikawa, T. Suzuki, H. Takenaga, K. Toi, and the JT-60 Team, "Energy loss for grassy ELMs and effects of plasma rotation on the ELM characteristics in JT-60U," *Nuclear Fusion*, vol. 45, no. 8, pp. 871–881, 2005.

[178] T. Osborne, J. Ferron, R. Groebner, L.L.Lao, A.W.Leonard, M. Mahdavi, R.Maingi, R. Miller, A. Turnbull, M.Wade, and J.Watkins, "The effect of plasma shape on H-mode pedestal characteristics on DIII-D," *Plasma Physics and Controlled Fusion*, vol. 42, no. Supplement 5A, pp. A175–A184, 2000.

[179] H.Zohm, *Magnetohydrodynamic Stability of Tokamaks*. Wiley-Vch, 2015.

This work was supported by the European Commission and carried out within the framework of the Erasmus Mundus International Doctoral College in Fusion Science and Engineering (FUSION-DC)